# Fallgruppen in der Versorgung von Patienten mit psychischen Störungen
# –
# Überprüfung der Eignung eines Fremdeinschätzungsinstrumentes „Die Health of the Nation Outcome Scales, HoNOS-D“ zur differenzierten Erfassung des Schweregrades im Rahmen der Entwicklung eines Klassifikationssystems

Dissertation zur Erlangung der Würde des Doktors der Philosophie

der Universität Hamburg

vorgelegt von


Sylke Andreas

aus Magdeburg


Hamburg, 2005

Referent: Prof. Dr. Dr. Uwe Koch

Korreferent: Prof. Dr. Franz Petermann

Datum der letzten mündlichen Prüfung: 21.September 2005

**Danksagung**

Ohne die Unterstützung, die ich von verschiedenen Seiten erhalten habe, wäre die vorliegende Arbeit nicht zustande gekommen. Ich möchte mich ganz herzlich bei Herrn Dr. Holger Schulz für seine intensive fachliche Unterstützung und zahlreiche Diskussionen und bei Herrn Prof. Dr. Dr. Uwe Koch für die Schaffung der Rahmenbedingungen und seine strategischen Impulse bedanken.

Ebenso möchte ich herzlich den Mitgliedern der Arbeitsgruppe für Psychotherapie- und Versorgungsforschung (AGPV) am Universitätsklinikum Hamburg-Eppendorf für zahlreiche fachliche Diskussionen danken.

Die vorliegende Arbeit ist im Rahmen eines Projektes des Norddeutschen Forschungsverbundes für Rehabilitationsforschung entstanden, gefördert durch das Bundesministerium für Bildung und Forschung sowie dem Verband Deutscher Rentenversicherungsträger und seinen Mitgliedern, denen ich an dieser Stelle auch meinen Dank aussprechen möchte. Maßgebliche Teile der vorliegenden Arbeit entstanden im Rahmen des Qualitätssicherungsprogramms der Gesetzlichen Krankenkassen. An dieser Stelle möchte ich ganz herzlich Dipl.-Psych. Timo Harfst und Dipl.-Psych. Stephan Kawski für die Projektkoordination danken.

Ich möchte mich weiterhin bei meinen Eltern Herrn Günter Andreas und Frau Marlies Andreas bedanken, da sie mir Unterstützung in jeglicher Form haben zu Teil werden lassen und mir bei der Verwirklichung meiner beruflichen Ziele immer hilfreich zur Seite gestanden haben.

Weiterhin möchte ich gern meinen besonderen Dank Herrn Sven Rabung aussprechen, der mich während der Anfertigung der Arbeit sowohl emotional als auch fachlich begleitet und unterstützt hat.

Hamburg, im Juli 2005 Sylke Andreas

# Inhaltsverzeichnis

# 1 Einleitung

In den letzten Jahren wurden in Deutschland unter ökonomischen Gesichtspunkten verstärkt neue Vergütungsformen im Krankenhaussektor diskutiert. Als eine Möglichkeit der Erstattung von Leistungen rückten hierbei **pauschalierte Entgeltsysteme** oder auch **„Diagnosis-Related-Group (DRG) Systems“** in den Mittelpunkt des Interesses. Mit der Verabschiedung des Gesundheitsreformgesetzes wurde im Dezember 1999 die Einführung eines solchen Systems in der Akutversorgung beschlossen. Damit sind mehrere **Zielsetzungen** verbunden. DRG-Systeme ermöglichen z.B. eine prospektive anreizorientierte Vergütung. Sie können außerdem zum Benchmarking zwischen Krankenhäusern beitragen oder zum internen Ressourcenmanagement eingesetzt werden. In diesem Kontext wird aber auch kritisch diskutiert, dass solche Systeme durchaus zu Lasten der Behandlungsgüte gehen könnten, weswegen gleichzeitig mit der Etablierung des DRG-Systems eine Einführung qualitätssichernder Maßnahmen verbunden sein sollte.

Pauschalierte Entgeltsysteme existieren bereits seit mehr als 20 Jahren und gehen auf **Entwicklungen von Fetter et al.** (1980) in den USA zurück. Auf der Basis von 500.000 Patientenangaben aus insgesamt 118 Institutionen entwickelten die Autoren das erste DRG-System für die Akutversorgung (somatischer und psychischer Bereich). Bereits drei Jahre später fand dieses Konzept im Rahmen der amerikanischen Krankenversicherung „Medicare“ Eingang in die Vergütung von Krankenhausleistungen. Mittlerweile sind DRG-Systeme weltweit in verschiedenen Ländern im somatischen Krankenhaussektor etabliert und ständig weiterentwickelt worden (vgl. hierzu Kapitel 2).

Das DRG-System in der von Fetter et al. (1980) konzipierten Fassung erwies sich in der Folgezeit aber *nicht für alle Versorgungsbereiche des Gesundheitswesens* als adäquat. Bereits 1984 wurde in der Arbeit von Taube et al. (1984b) festgestellt, dass die Kategorien so wie sie vom **DRG-System für den Bereich psychischer Störungen** vorgehalten werden, die Behandlungsdauer bzw. den Ressourcenverbrauch aufgrund verschiedener Gründe z.B. unterschiedliche Therapieverläufe trotz gleicher Hauptdiagnose von Patienten mit psychischen Störungen nicht ausreichend abbilden können. Deswegen wurde der stationäre Versorgungsbereich psychisch Kranker auch **in anderen Ländern** (u.a. USA, Neuseeland, Australien) von der DRG-basierten Vergütung ausgeschlossen. Betrachtet man nun die Entwicklungen in den Ländern so lässt sich feststellen, dass dieser Ausschluss einige Nachteile nach sich zog. Im Bereich psychischer Störungen kam es u.a. zu einer defizitären

Entwicklung im Hinblick auf die Einführung qualitätssichernder Maßnahmen im Vergleich zur somatischen Medizin (Buckingham et al., 1998). Auch aus diesem Grund gibt es seit Anfang der 80-er Jahre weltweit Bestrebungen, für den psychischen Versorgungsbereich *adäquatere Konzepte zur Fallgruppierung* zu entwickeln. Die Ergebnisse dieser Bemühungen sind in Kapitel 3 dieser Arbeit ausführlich dargelegt.

**In Deutschland** wurde der Bereich psychischer Störungen ebenfalls aufgrund zahlreicher empirischer Befunde und nach Expertenmeinungen zunächst von der Einführung des neuen pauschalierten Entgeltsystems ausgeklammert. Dazu gehen die Meinungen allerdings weit auseinander, was in den unterschiedlichen Positionen der vertretenden Fachgesellschaften (wie z.B. des Deutschen Kollegiums für Psychosomatische Medizin DKPM oder der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Neurologie DGPPN) zum Ausdruck kommt. Aufgrund dieser mittel- bzw. langfristig noch unsicheren Perspektive und der internationalen Erfahrungen bedarf es der **differenzierten Auseinandersetzung mit Klassifikationssystemen für den Bereich psychischer Störungen**.

Während allgemeine DRG-Systeme von einer zentralen Stellung der Hauptdiagnose als Klassifikationskriterium ausgehen, betonen zahlreiche Autoren die **Bedeutung der differenzierten Erfassung des Schweregrades** einer Erkrankung für die Fallgruppenbildung psychisch Kranker (Mezzich, 1991). Derzeit fehlt es in Deutschland auf nationaler Ebene sowohl an einem adäquaten Ansatz zur Klassifikation von Patienten[1] mit psychischen Störungen als auch an einem Instrument, welches über verschiedene Versorgungsbereiche psychisch Kranker hinweg geeignet wäre, die *Schwere einer psychischen Störung* differenziert zu erfassen (vgl. hierzu auch Kap. 5 und Kap. 6).

Internationale komplementär zu den DRG-Systemen entwickelte Fallgruppenkonzepte beruhen überwiegend auf **Fremdeinschätzungen**. Aufgrund der Problematik, den Schweregrad bei stärker belasteten Patienten adäquat erfassen zu können, wurde deswegen als

---

[1] Aufgrund besserer Lesbarkeit wird im Folgenden durchgängig die grammatikalisch männliche Form bei Patienten, Behandlern usw. benutzt. Darin eingeschlossen ist aber jeweils auch die weibliche Form.

**zentrales Element** des Klassifikationssystems ein Fremdeinschätzungsinstrument zur differenzierten Erfassung des Schweregrades gesucht. Auf der Basis internationaler empirischer Befunde kristallisierte sich insbesondere ein Instrument als bedeutsamer Prädiktor des Ressourcenverbrauchs heraus, die **„Health of the Nation Outcome Scales, HoNOS“**. Im Jahr 2003 wurde die HoNOS, autorisiert durch die Testautoren (Royal College of Psychiatrists´), im Rahmen eines mehrstufigen, experten-gestützten Konsensusprozesses von der Arbeitsgruppe für Psychotherapie- und Versorgungsforschung (AGPV) des Instituts für Medizinische Psychologie am Universitätsklinikum Hamburg-Eppendorf in die deutsche Sprache übersetzt (HoNOS-D) (Andreas et al., 2004b).

Die HoNOS wurde in Großbritannien u.a. unter der Prämisse entwickelt, *in allen Versorgungsbereichen* (z.B. stationär, ambulant) psychisch Kranker einsetzbar zu sein. Damit ist nun zum einen die Voraussetzung gegeben, ein Klassifikationssystem zu entwickeln, welches in verschiedenen Versorgungsbereichen der stationären Behandlung psychisch Kranker einsetzbar sein könnte und zum anderen die in der vorliegenden Untersuchung gewonnenen Ergebnisse auf eine international diskutierbare Plattform zu stellen. In der vorliegenden Arbeit soll nun, eingebettet in die Entwicklung eines Patientenklassifikationssystems für den Bereich psychischer Störungen in Deutschland, die Eignung der deutschen Version der HoNOS (HoNOS-D) überprüft werden.

Zur Skizzierung des Untersuchungsrahmens der vorliegenden Arbeit werden im Folgenden zunächst die Charakteristika des deutschen Versorgungssystems für Patienten mit psychischen Störungen vorgestellt. Dabei soll beschrieben werden, wie sich das System in der Bundesrepublik Deutschland herausgebildet hat und welche Bereiche an der stationären Versorgung von Patienten mit psychischen Störungen beteiligt sind.

Das Gesundheitssystem der Bundesrepublik Deutschland verfügt über ein im internationalen Vergleich außergewöhnlich intensiv ausgebautes stationäres System der psychotherapeutischen Versorgung. Diese stationären Angebote verteilen sich, wenn auch ungleichgewichtig, auf drei unterschiedliche Versorgungsbereiche, nämlich auf die psychosomatisch-psychotherapeutische Krankenhausbehandlung, die psychiatrisch-psychotherapeutische Krankenhausversorgung und die Rehabilitation von Patienten mit psychischen Störungen (Schulz & Koch, 2002). Strukturell gesehen sind an dieser Stelle zwei weitere Besonderheiten dieses Systems zu nennen: (1) Ein substanzieller Teil des stationären Versorgungs-

systems hat sich nicht außerhalb (wie international üblich), sondern innerhalb des psychiatrisch-psychotherapeutischen Versorgungssystems entwickelt und (2) die Zahl der Betten ist überwiegend in der medizinischen Rehabilitation und nicht in der akuten Krankenhausversorgung angesiedelt. In einer Expertise zur stationären Versorgung psychisch Kranker in Norddeutschland führen Schulz und Koch (2002) auf, dass im Jahr 1993 72% der psychotherapeutischen Betten auf die medizinische Rehabilitation und 28% der Betten auf die Krankenhausbehandlung entfielen. Historisch betrachtet liegen die Gründe für diese besondere Entwicklung hauptsächlich in der starken biologischen Ausrichtung der Psychiatrie nach dem Krieg, einer Reduktion von Betten in psychiatrischen Landeskrankenhäusern und einer Spezialisierung auf die psychotherapeutische Behandlung psychischer Störungen im Rahmen der medizinischen Rehabilitation (Koch & Potreck-Rose, 1994).

Die **stationäre Krankenhausbehandlung** wird von Krankenhäusern und Abteilungen für psychotherapeutische Medizin und Psychosomatik bzw. von Krankenhäusern für Psychiatrie und Psychotherapie durchgeführt. Kostenträger sind im Wesentlichen die gesetzlichen Krankenkassen, zum Teil erfolgt die Erstattung der Leistungen auch durch Privat- und Selbstzahlungen. Laut der Gesundheitsberichterstattung von Schulz et al. (in Druck) gab es im Jahr 2002 81 **Fachabteilungen für Psychosomatik und Psychotherapie** mit insgesamt 2.996 Betten und einer Behandlungszahl von 24.418 Patienten mit psychischen Störungen. Am häufigsten werden in diesem Bereich Patienten mit einer affektiven Störung (37%) behandelt, gefolgt von Patienten mit einer Neurotischen-, Belastungs- und somatoformen Störung (24%). Die durchschnittliche Verweildauer betrug 43,3 Tage. Im Jahr 2002 gab es 396 **Fachabteilungen für Psychiatrie und Psychotherapie** mit 53.915 Betten und einer Behandlungsanzahl von 632.493 Patienten. Die durchschnittliche Verweildauer betrug 27,5 Tage. Häufigste Erstdiagnosen in diesem Versorgungsbereich sind Störungen im Zusammenhang mit psychotropen Substanzen (35%), Affektive Störungen (20%) und Schizophrenie (20%) (Schulz et al., in Druck).

Ein zweiter Sektor der stationären Versorgung psychisch Kranker wird von der **Rehabilitationsbehandlung von Patienten mit psychosomatischen/psychotherapeutischen Störungen** abgedeckt. In diesem Fall sind die Kostenträger im Wesentlichen die gesetzliche Rentenversicherung, zum Teil erfolgt die Erstattung der Leistungen auch durch die gesetzliche Krankenversicherung. Im Jahr 2001 gab es hier 175 Fachabteilungen für Psychosomatik und Psychotherapie mit insgesamt 15.421 Betten. Im Jahr 2002 wurden nach Schulz

et al. (in Druck) 140.460 Patienten in der stationären Rehabilitation psychosomatisch/psychischer Störungen mit einer durchschnittlichen Liegedauer von 38,2 Tagen behandelt. Die häufigste „Erst-Diagnose“ waren hier psychische Störungen durch psychotrope Substanzen (28,6%) gefolgt von Affektiven Störungen (25,6%) und Reaktionen auf schwere Belastungen bzw. Anpassungsstörungen (17,2%).

Die **Behandlung der Patienten mit psychischen Störungen** in den oben aufgeführten Sektoren erfolgt zumeist multimodal unter Beteiligung verschiedener Berufsgruppen (z.B. Ärzte, Psychologen, Pflegekräfte) und multifaktoriell hinsichtlich unterschiedlicher Therapieangebote (z.B. Einzel- und Gruppentherapie mit ergänzenden Körper- und Kreativtherapieangeboten).

Seit mehreren Jahren wird immer wieder kritisch diskutiert, in welchen **Faktoren** sich die **stationäre Krankenhausbehandlung von der Rehabilitationsbehandlung** für Patienten mit psychosomatischen/psychotherapeutischen Störungen **unterscheidet**. Dabei können Kriterien, wie sie für die somatische Medizin gelten (z.B. Maßnahmen zur Reintegration psychisch Kranker) nicht unmittelbar auf den psychischen Versorgungsbereich angewandt werden. Schulz und Koch (2002) diskutieren im Rahmen der Expertise zur stationären Versorgung psychisch Kranker in Norddeutschland verschiedene Kriterien zur Abgrenzung dieser beiden Versorgungsbereiche. Dabei wurde deutlich, dass bereits auf der gesetzlichen Ebene (§ 107 SGB V) eine Abgrenzung schwer fällt. So fassten Schulz und Koch (2002) zusammen, dass der Unterschied allenfalls darin besteht, dass die Krankenhausbehandlung zusätzlich diagnostische Aufgaben beinhaltet und die Rehabilitationsbehandlung von Patienten mit psychischen Störungen darüber hinaus explizit den erzielten Behandlungserfolg sichern bzw. Pflegebedürftigkeit verhindern sollte. Weiterhin zeigten Schulz und Koch (2002) auf, dass hinsichtlich der Behandlungsmaßnahmen eine etwas klarere Abgrenzung gelingt. Nach dem Gesetzestext erfolgt die stationäre Krankenhausbehandlung in höherem Ausmaß durch ärztlich-pflegerische Hilfsleistungen; die Rehabilitationsbehandlung hingegen häufiger durch die Anwendung von Heilmitteln sowie anderer geeigneter Maßnahmen.

Im Rahmen der Expertise wurde auch deutlich, dass der Gesetzestext des Paragraphen 107 SGB V nicht die Charakteristika der Indikationsbereiche berücksichtigt. Daraus schlussfolgern Schulz und Koch (2002), dass der Gesetzestext für die Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie fachlich nicht angemessen ist. Weiterhin füh-

ren die Autoren auf, dass in der Fachliteratur häufig folgende Kriterien zur Abgrenzung der beiden Bereiche genannt werden: *Zielsetzung der Behandlung*, *Zeitpunkt des Behandlungsbeginns*, *Ausmaß vitaler Gefährdung*, *Schwere der Alltagsbeeinträchtigung*, *Art und Dichte der diagnostischen und therapeutischen Maßnahmen*, *erforderlicher pflegerischer*, *apparativer Aufwand* und *Durchführung der Initialdiagnostik*. Die Autoren ziehen nach Überprüfung dieser Merkmale das Fazit, dass die aufgeführten Kriterien allenfalls eine relative Diskriminationsfähigkeit zwischen Krankenhaus- und Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen zulassen.

**Zusammenfassend** und der Argumentation von Schulz und Koch (2002) folgend, besteht sicherlich ein hoher Überschneidungsbereich zwischen den stationären Versorgungsbereichen für Patienten mit psychischen Störungen. Daraus lässt sich ableiten, dass Entwicklungen für den Akutbereich auch für den Bereich der Rehabilitationsbehandlung psychisch Kranker relevant sein könnten und umgekehrt. Wenn also die Einführung eines Klassifikationssystems aus Sicht der Fachverbände und der Kostenträger in der Akutversorgung und/oder Rehabilitationsbehandlung psychisch Kranker vorstellbar wird, sollte bereits ein adäquates Konzept vorgehalten werden können. Vorliegende Untersuchung fokussiert auf den Bereich der Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen. Im Folgenden wird näher auf den Aufbau der Arbeit hingewiesen.

Kapitel 2 gibt einen Überblick über den Stand der Einführung des DRG-Systems in Deutschland. Anschließend sollen im Rahmen einer Literaturanalyse internationale komplementär zu den DRG-Systemen entwickelte Klassifikationskonzepte vorgestellt und zentrale Parameter der Fallgruppenbildung abgeleitet werden (Kap. 3). Im Kapitel 4 werden die international gefundenen Dimensionen und Kriterien um nationale empirische Befunde ergänzt und auf zentrale Bestandteile zur Gruppierung überprüft. Kapitel 5 gibt einen Überblick über bereits im deutschen Sprachraum etablierte Selbst- und Fremdeinschätzungsinstrumente zur Erfassung der Schwere einer psychischen Erkrankung. Die Entwicklung der HoNOS und empirische Befunde zu psychometrischen Eigenschaften werden in Kapitel 6 referiert. In Kapitel 7 soll der Übersetzungsprozess der deutschen Version der HoNOS vorgestellt und dessen Aufbau skizziert werden. Nach der Ableitung der Hypothesen und Fragestellungen (Kap. 8), wird im Kapitel 9 das methodische Vorgehen der Unter-

suchung beschrieben. Im Kapitel 10 werden die Ergebnisse präsentiert und abschließend in kritisch diskutiert. Kapitel 11 vermittelt eine zusammenfassende Darstellung.

## 2 Fallgruppen in der Akutmedizin

Bislang wurden erbrachte Leistungen in deutschen Krankenhäusern über Abteilungs- und Basispflegesätze und Sonderentgelte vergütet. Mit Inkrafttreten des Gesundheitsreformgesetzes ändert sich ab 2003 die Abrechnungssituation der Krankenhäuser, ein neues Entgeltsystem auf der Basis von Fallpauschalen wird im Bereich der Akutversorgung eingeführt. Das folgende Kapitel soll einen Überblick über das neue Entgeltsystem, den Diagnosis-Related-Groups (DRGs) und den damit verbundenen Veränderungen geben.

### 2.1 Ziele und Funktionen von DRG-Systemen

Mit der Einführung eines pauschalierten Entgeltsystems in Deutschland werden vielfältige Zielsetzungen verfolgt. Zum einen wird eine leistungsorientierte Vergütung der Krankenhäuser mit Erhöhung der Transparenz über Leistungs- und Kostenstrukturen und zum anderen eine effizientere Nutzung von Wirtschaftlichkeitsressourcen mit Anreizsetzung zur Erhöhung des Wettbewerbes um eine qualitativ gute Versorgung angestrebt (Tuschen, 2001). Ein weiteres Ziel besteht in der Verkürzung der im internationalen Vergleich zu langen Behandlungsdauern von Patienten im Krankenhaus. Oberste Prämissen sind die Optimierung der Versorgungsstrukturen und gleichzeitig eine Stabilisierung der Ausgaben der gesetzlichen Krankenversicherung (Tuschen, 2001).

In verschiedenen Ländern in Europa kommen DRG-Systemen unterschiedliche Funktionen zu. Wie aus Abbildung 1 ersichtlich dienen sie z.B. in den skandinavischen Ländern zur Abrechnungssteuerung, wohingegen sie z.B. in Frankreich als Mittel zur Budgetbemessung eingesetzt werden. D.h. dass die aufgrund des DRG-Systems gewonnenen Leistungsdaten im Rahmen der prospektiven Festlegung des Jahresbudgets herangezogen werden, aber nicht zur fortlaufenden Abrechnung dienen (Rochell & Roeder, 2000). In Belgien und Irland werden DRG-Systeme zur Steuerung der Verweildauern eingesetzt. Das bedeutet, dass für die Krankenhäuser Über- oder Unterschreitungen der landesweiten Durchschnittsverweildauern pro Fallgruppe relevant und mit Konsequenzen verbunden sind (u.a. Positivanreize, Abzüge im Negativfall). Als Mittel zur Abrechnung der stationären Krankenhausleistungen werden DRGs landesweit oder regional mittlerweile in acht europäischen Ländern eingesetzt (z.B. Norwegen, Finnland, Schweden; Rochell & Roeder, 2000; Roeder, 2000).

Betrachtet man die Entwicklung außerhalb von Europa, so lässt sich festhalten, dass z.B. in Australien und in den Vereinigten Staaten Amerikas das DRG-System zum Controlling, Benchmarking und zur Abrechnung über Fallpauschalen verwendet wird (Fischer, 2003).

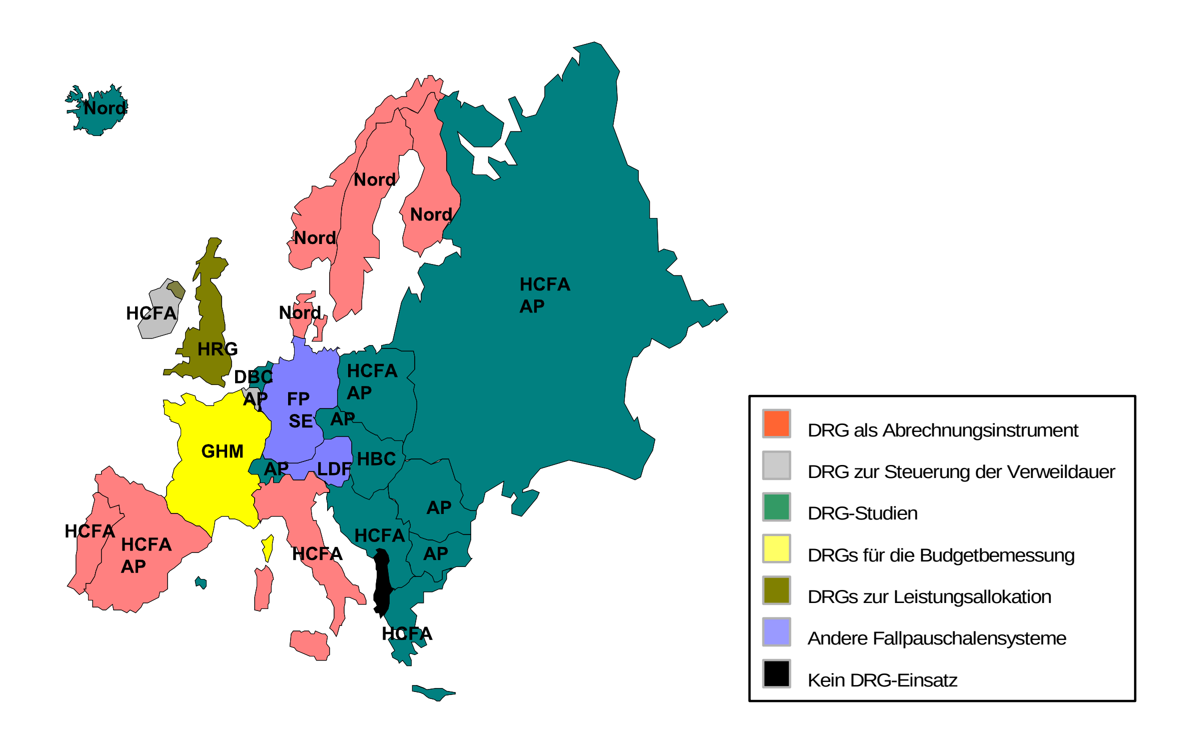



Abbildung 1: Funktionen von international bereits etablierten DRG-Systemen (Diagnosis-Related-Groups, DRG)
(Darstellung aus Rochell, 2000 übernommen)

## 2.2 Ausgangslage in Deutschland

Durch die Verabschiedung des Gesundheitsreformgesetzes am 22.12.1999 wurde die Einführung eines neuen Entgeltsystems im Krankenhaussektor deutschlandweit vorbereitet. Mit Inkrafttreten des § 17b Abs. 2 des Krankenhausfinanzierungsgesetzes wurden die Krankenhäuser dazu verpflichtet, ein durchgängiges, leistungsorientiertes und pauschalierendes Vergütungssystem einzuführen, das sich an einem international bereits eingesetzten Fallpauschalensystem orientiert.

Anstatt der Vergütung über Abteilungs- und Basispflegesätze, Sonderentgelte und Fallpauschalen, die neben der Operationsleistung auch die Behandlungsleistung und die Unter-

bringungskosten vergüten (Clade, 2000; Egner & Verbarg, 2001), soll zukünftig die Abrechnung über die Ermittlung eines Pauschalentgeltes anhand der DRG-Fallgruppe (Commonwealth Department of Health and Family Services, 1998b) mit einem für jede DRG spezifischen Preis erfolgen (Rochell & Roeder, 2002). Das Prinzip der DRGs beruht hierbei auf der Zuordnung von Patienten zu Fallgruppen aufgrund festgelegter Kriterien (z.B. Hauptdiagnose), die sich in ihrem Ressourcenverbrauch möglichst gleichen und äquivalente Kosten verursachen.

Als Basis zur **Entscheidungsfindung**, welches DRG-System an deutsche Verhältnisse adaptiert werden soll, wurde von der Deutschen Krankenhausgesellschaft zum einen ein **gutachterlicher Systemvergleich** in Auftrag gegeben (Fischer, 2002) und zum anderen ein **Vergleich verschiedener DRG-Systeme** in Kooperation mit der Deutschen Gesellschaft für Thorax-, Herz- und Gefäßchirurgie und der Westfälischen Wilhelms-Universität Münster durchgeführt (Roeder, 2000).

In dem von Fischer erstellten **Gutachten** wurden zehn verschiedene, international bereits etablierte DRG-Systeme (u.a. die amerikanischen Diagnosis Related Groups der Health Care Financing Administration, HCFA-DRG oder das französische System Groupes homogénes de malades, GHM) hinsichtlich ihrer Entwicklungspotentiale und Grundprinzipien miteinander verglichen und anschließend bewertet (Bundesministerium für Gesundheit, 2002). Kriterien, die der Begutachtung zugrunde lagen, waren u.a. eine Differenzierungsfähigkeit nach Schweregraden anhand der vorgesehenen Komplikationsstufen oder die Bewertung der abhängigen Variable. Es folgte eine Analyse der Grundzüge von DRG-Systemen und ihrer Konstruktionsansätze. Zusammenfassend kam Fischer (Bundesministerium für Gesundheit, 2002) zu dem Schluss, dass das australische Fallgruppensystem (Australian Refined Diagnosis Related Group System, AR-DRG System) am ehesten als Grundlage für eine deutsche Adaptation geeignet erscheint. „Wenn ich gleich heute ein System vorschlagen müsste, ..., dann würde ich als Erstes das AR-DRG-System aus Australien in Betracht ziehen, da es gut formalisiert und dadurch - trotz seiner Komplexität - einigermaßen transparent und relativ flexibel ist.“ (zitiert nach Kapitel G, S. 2 Bundesministerium für Gesundheit, 2002).

Roeder und Rochell (2000) verglichen anhand eines Evaluationskataloges acht verschiedene DRG-Systeme auf der Basis von Krankenhausdaten. Begutachtet wurden u.a. die

Australian Refined Diagnoses Related Groups, AR-DRG oder die Health Care Financing Administration, HCFA-DRG´s. Datengrundlage bildeten Aufnahme- bzw. Entlassungsangaben sowie Kostenkalkulationen von Patienten aus der Thorax-, Herz- und Gefäßchirurgie. Die Autoren sprachen zwar keine konkreten Empfehlungen für eines der untersuchten Systeme aus, zeigten jedoch erhebliche Unterschiede zwischen den Systemen u.a. bezüglich der Transparenz des Gruppierungsverfahrens, des Umfangs und der Aussagekraft der Dokumentation und sachgerechter Abbildung der Schweregrade auf. Die Autoren kamen zu dem Schluss, dass das französische GHM-System und das australische AR-DRG-System im Bereich der Herzchirurgie den aktuellsten medizinischen Bezug aufwiesen.

Gestützt auf die positive Bewertung durch den gutachterlichen Systemvergleich von Fischer (Bundesministerium für Gesundheit, 2002) und den DRG-Praxisvergleich von Roeder und Rochell (Roeder & Rochell, 2000) sowie sich diesem Votum anschließenden Empfehlungen der Bundesärztekammer und des Deutschen Pflegerates fiel die Entscheidung der Selbstverwaltung der Spitzenverbände der Krankenkassen (GKV), des Verbandes der privaten Krankenversicherung (PKV) und der Deutschen Krankenhausgesellschaft (DKG) auf das **australische AR-DRG-System** (Version 4.1).

### 2.3 Das australische Fallgruppensystem (AR-DRG-System)

Das australische AR-DRG-System 4.1 wurde 1998 nach langjähriger Entwicklungs- und Überarbeitungszeit veröffentlicht und löste das seit 1993 existierende australische AN-DRG-System (Australian National DRGs) ab. Zur Überarbeitung des vorliegenden Systems wurden 20 Expertengruppen gebildet, welche die Schweregradeinteilung und u.a. auch einige Hauptkategorien intensiv modifizierten und verfeinerten (Refinement-Ansatz) (Commonwealth Department of Health and Family Services, 1998a).

Das Prinzip der Zuordnung von Patienten zu einer Fallgruppe beruht auf einem hierarchischen Gruppierungsalgorithmus, wobei jeder Behandlungsfall ausgehend von den Entlassungsdaten exakt einer DRG-Fallgruppe zugeordnet wird (Fischer, 2002; Rochell & Roeder, 2002). Im AR-DRG-System sind dies insgesamt 661 Fallgruppen, die zu 23 Hauptkategorien (MDC, Major Diagnostic Categories) gehören (Commonwealth Department of Health and Family Services, 1998a). Tabelle 1 bietet eine Darstellung über alle im AR-

DRG-System vorkommenden Hauptkategorien. Anschließend soll der Gruppierungsprozess beschrieben werden.

Tabelle 1: Die Hauptkategorien des AR-DRG-Systems (Fischer, 2001)

| MDC | Bezeichnung | MDC | Bezeichnung |
|---|---|---|---|
| **00** | Sonderfälle/ Pre MDC | **12** | Erkrankungen der männlichen Geschlechtsorgane |
| **01** | Erkrankungen des Nervensystems | **13** | Erkrankungen der weiblichen Geschlechtsorgane |
| **02** | Erkrankungen des Auges | **14** | Schwangerschaft und Geburt |
| **03** | Erkrankungen des Hals-, Nasen-, Ohrenbereiches | **15** | Neugeborene |
| **04** | Erkrankungen der Atemwegsorgane | **16** | Erkrankungen des Blutes und blutbildender Organe |
| **05** | Erkrankungen des Kreislaufsystems | **17** | Neubildungen (hämatologische und solide) |
| **06** | Erkrankungen des Verdauungssystems | **18** | Infektionen und parasitäre Erkrankungen |
| **07** | Erkrankungen der Leber, Galle, Pankreas | **19** | **Psychische Störungen** |
| **08** | Erkrankungen der Muskeln, des Skelettes und des Bindegewebes | **20** | **Alkohol- und Drogenabusus, Alkohol- und Drogeninduzierte organische psychische Störungen** |
| **09** | Erkrankungen der Haut, des subcutan Gewebes und der Mamma | **21** | Verletzungen und Vergiftungen |
| **10** | Hormonelle und Stoffwechselerkrankungen | **22** | Verbrennungen |
| **11** | Erkrankungen der Nieren und Harnwege | **23** | Andere Faktoren, die den Gesundheitszustand beeinflussen |

**Anmerkungen:** MDC = Major Diagnostic Category (Hauptkategorie des DRG-Systems); fett hervorgehoben = Gruppierungskategorien für den psychischen Bereich

Zur besseren Veranschaulichung ist in Abbildung 2 der **Gruppierungsprozess des australischen Fallgruppensystems** graphisch dargestellt. Dabei werden zunächst nicht zuzuord-

nende oder besonders aufwändige Fälle (Ausnahmefälle) identifiziert und in einer eigenen Kategorie zusammengefasst. Die verbleibenden Behandlungsfälle werden auf der Basis der ICD-10-Kodierung[2] einer von 23 Hauptkategorien (Major Diagnostic Categories, MDCs) zugeteilt (siehe Tabelle 1). Auf dieser Ebene werden die Fälle, je nachdem welche medizinischen Leistungen (z.B. Operation) erbracht wurden, in drei weitere Gruppen (Sub-MDCs) gegliedert (operative, medizinische und andere Unterkategorien) (siehe Abbildung 2). Innerhalb der Gruppen werden dann vergleichbare Fälle anhand der Hauptdiagnose und weiterer Kriterien (z.B. Alter) in eine von insgesamt 409 Basis-DRGs aufgeteilt. Die Nebendiagnose, d.h. die Ko- oder Multimorbidität, dient hierbei der Ermittlung eines kumulierten Schweregrades, der fünf Ausprägungen[3] annehmen kann (Abbildung 2):

CCL = 0: Nebendiagnose ist keine Komplikation oder Komorbidität
CCL = 1: leichte Komplikation/Komorbidität (minor)
CCL = 2: mittleres Komplikations-/Komorbiditätsniveau (moderate)
CCL = 3: schwereres Komplikations-/Komorbiditätsniveau (severe)
CCL = 4: sehr schweres Komplikations-/Komorbiditätsniveau (catastrophic)

Das bedeutet, dass jeder Nebendiagnose eine Komplikations- oder Komorbiditätsstufe zugeordnet wird. Verschiedene Stufen können anschließend in einem Patient Clinical Complexity Level (PCCL) zusammengefasst werden (vgl. Abbildung 2; Commonwealth Department of Health and Family Services, 1998a).

Jeder der insgesamt 661 Fallgruppen (AR-DRGs) wird nun ein relatives Kostengewicht zugeordnet, dessen Produkt mit dem vorgegebenen Basisfallpreis die Ermittlung eines Pauschalentgelts ermöglicht (siehe Abbildung 2).

---

[2] ICD-10 Australian Modification auf ICD-10 basierend (Diagnosecodes; Fischer, 1997)

[3] bei chirurgischen Fällen Skala 0-4, bei medizinischen Fällen Skala 0-3

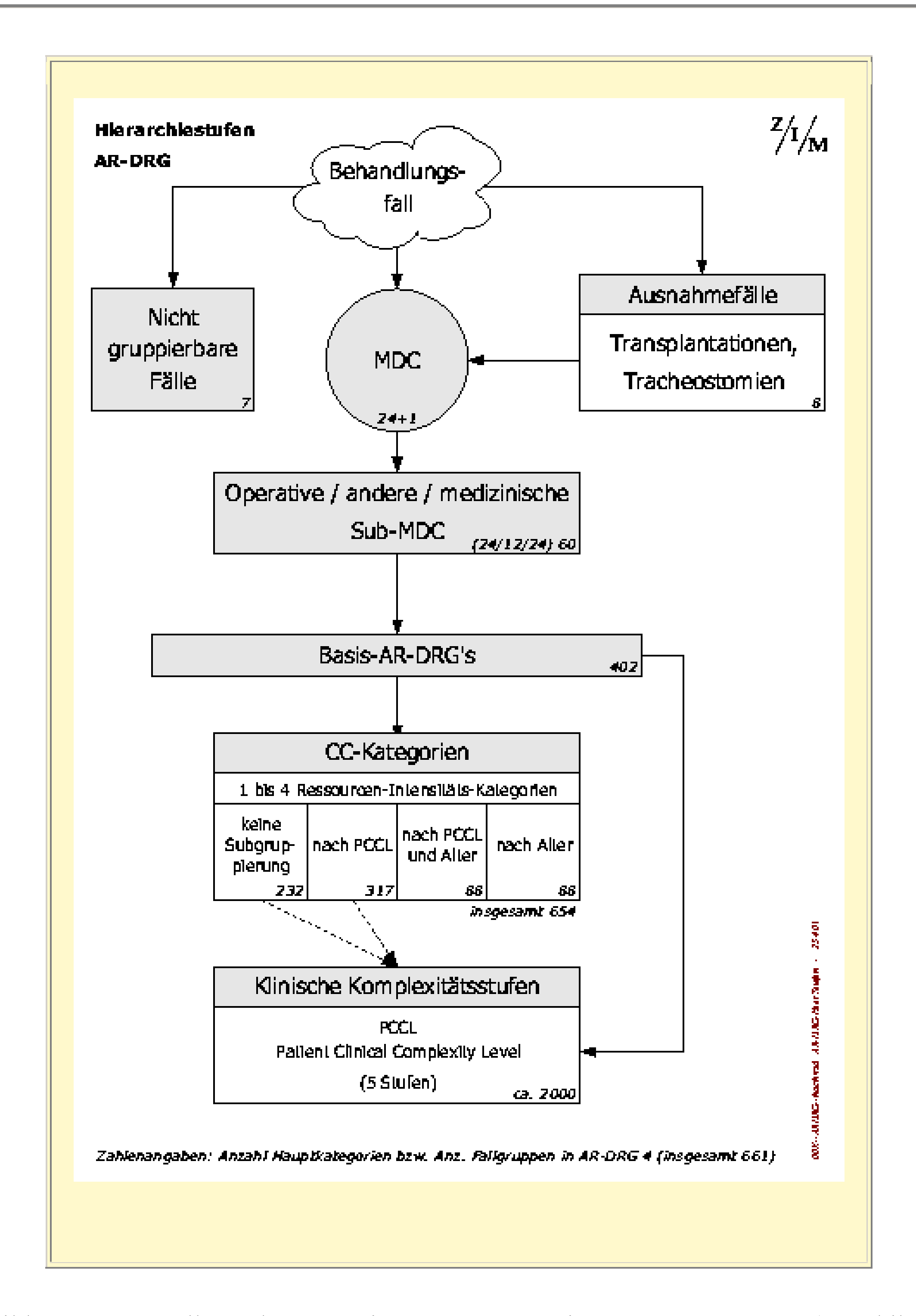



Abbildung 2: Darstellung des Gruppierungsprozesses im AR-DRG-System (Graphik entnommen aus Fischer, 2000). Anmerkungen: AR-DRG = Australian Refined Diagnosis Related Groups; MDC = Major Diagnostic Category (Hauptkategorie des DRG-Systems); Sub-MDC = Untergruppe der Major Diagnostic Category; CC = Komorbidität oder Komplikation (Comorbidity or Complication); PCCL = Patient Clinical Complexity Level.

### 2.4 Bewertung des AR-DRG-Systems

Patientenklassifikationssysteme, so wie auch das AR-DRG-System, wurden mit dem Ziel entwickelt, bezüglich des Ressourcenverbrauchs *homogene* Behandlungsfallgruppen zu definieren. Damit wird angestrebt, dass die Streuung der Kosten, der in einer Gruppe zusammengefassten Behandlungsfälle, relativ gering sein soll.

Um die Homogenität in einer Fallgruppe in allgemeinen DRG-Systemen zu erhöhen, wurden verschiedene Maßnahmen ergriffen: 1) Zunehmende Verfeinerung der DRGs durch Differenzierung der Komorbiditäten, 2) Zunehmende Differenzierung durch Einführung verschiedener Alterssplits und 3) Verwendung von zusätzlichen Gruppierungskriterien (wie z.B. Geburtsgewicht oder Notaufnahme) (Fischer, 2000).

Obwohl das AR-DRG-System über einen relativ differenzierten Klassifikationsalgorithmus unter Berücksichtigung von Komorbiditäten bzw. Komplikationen verfügt (siehe Abbildung 2) und deswegen zur Adaption an deutsche Verhältnisse aufgrund eines gutachterlichen Systemsvergleichs von Fischer empfohlen wurde (Bundesministerium für Gesundheit, 2002), erzielt es im Vergleich zu anderen DRG-Systemen kaum bessere Homogenitätskennwerte (Fischer, 2001).

Vorgelagerte Ursachen können eine nicht angemessene Kodierungsqualität oder nicht adäquate Kalkulationsmethoden von Soll- und/oder Ist-Kosten sein. Es kann sich hierbei aber auch um generelle Konstruktionsprobleme von DRG-Systemen handeln. So ist denkbar, dass unpräzise Falldefinitionen, eine Nicht-Berücksichtigung weiterer kostenrelevanter Parameter oder eine fehlende Kodierungsmöglichkeit für Mehrfachbehandlungen vorliegen könnten. An dieser Stelle besteht sicherlich für die Zukunft weiterer Überarbeitungs- und Optimierungsbedarf.

### 2.5 Der Adaptationsprozess an deutsche Verhältnisse

Das australische DRG-System soll nun auf der Basis bundesweit erhobener Daten an die Bedingungen in Deutschland adaptiert werden. Grundlage bilden zum einen dokumentierte Routinedaten der Patienten und zum anderen auf der strukturellen Ebene die Kostendaten aller Krankenhäuser. Die Datenerhebung soll retrospektiv erfolgen und sich auf ein abge-

schlossenes Kalenderjahr beziehen (Janiszewski & Pfaff, 2001; Irps et al., 2002). Von der Selbstverwaltung der GKV, der PKV und der DKG wurden Mitte April 2001 Kodierrichtlinien veröffentlicht, die eine an deutsche Verhältnisse angepasste Übersetzung des allgemeinen Teils der australischen Regeln darstellen. Diese sind seit 1. Januar 2002 verbindlich (Irps et al., 2002).

Vom Bundesministerium für Gesundheit und Soziales (BMGS) wurde 2001 ein Eckpunktepapier zur Einführung des deutschen DRG-Fallpauschalensystems veröffentlicht und zur Diskussion gestellt (Meu, 2001). Abbildung 3 veranschaulicht den Einführungsprozess im Zeitkontinuum. Demnach besteht ab dem 1. Januar 2003 für alle Krankenhäuser die Möglichkeit, auf das neue Vergütungssystem umzustellen (Tuschen et al., 2003). Für das Jahr 2004 ist dann eine verbindliche Anwendung des neuen Entgeltsystems vorgesehen (Egner & Verbarg, 2001). Allerdings gilt sowohl für das Jahr 2003 als auch für das Jahr 2004 eine budgetneutrale Phase (siehe Abbildung 3; Bundesministerium für Gesundheit, 2002). Anschließend sollen in der Konvergenzphase jeweils zum 1. Januar der Jahre 2005, 2006 und 2007 die Erlösbudgets der Krankenhäuser schrittweise an den festzulegenden landeseinheitlichen Basisfallwert und das sich daraus ergebende Erlösvolumen angeglichen werden. Wird dieser Zeitplan der gesetzlichen Vorgaben eingehalten, so würden die Krankenhäuser ab 2007 über landeseinheitliche Basisfallwerte ihre Fallpauschalen abrechnen (siehe Abbildung 3).

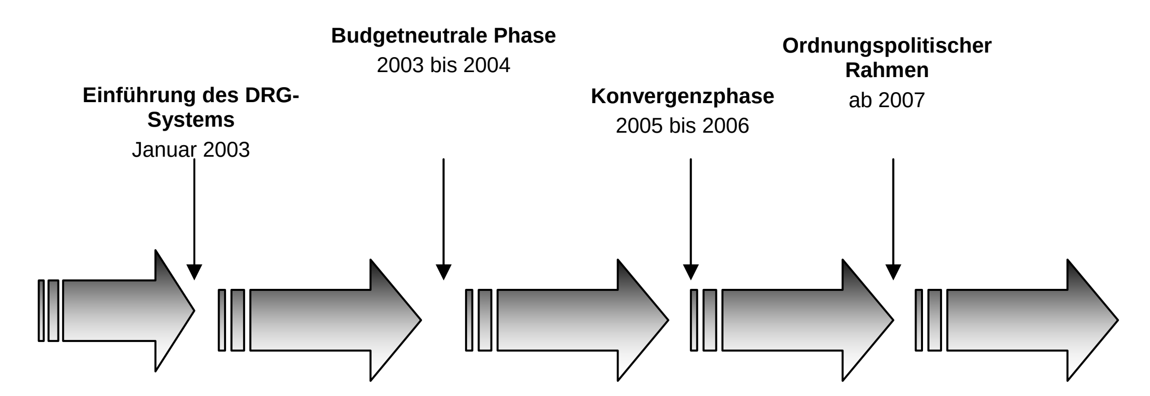



Abbildung 3: Zeitachse des Eckpunktepapier des BMGS (Bundesministerium für Gesundheit und Soziales) zur DRG-Einführung in Deutschland (Meu, 2001)

### 2.6 Erwartete Auswirkungen der DRG-Einführung

Wie langjährige internationale Erfahrungen nach der Einführung von Fallpauschalen zeigen, steht auch Deutschland mit der Einführung des DRG-Systems vor einem weitgreifenden Umstrukturierungsprozess im Leistungsabrechungssystem der Kostenträger (Haaf & Röckelein, 2002). Basierend auf Ergebnissen empirischer Studien waren im Wesentlichen folgende Entwicklungen für die Krankenhäuser weltweit zu beobachten (Rochell & Roeder, 2002):

- Signifikante Rückgänge der stationären Verweildauern
- Anstieg des Casemix-Index, d.h. der durchschnittlichen Fallschwere
- Steigerung der Produktivität der Krankenhäuser, d.h. Anstieg der Fallzahl bei gleichzeitiger Reduktion der Kosten je Fall und Erhöhung der Kosten je Behandlungstag durch eine höhere Leistungsdichte.

DRG-Systeme führen aber nicht nur zu Veränderungen auf der Krankenhausebene, sondern können auch zu einem Umstrukturierungsprozess des gesamten klinischen Versorgungssystems beitragen. Coffey (1999) zeigt in einer Übersichtsarbeit u.a. folgende Veränderungen seit der Einführung des Medicare Prospective Payment Systems (PPS) im Jahre 1983 in den USA auf der Gesamtstrukturebene auf:

- Anstieg von ambulant durchgeführten Operationen (1980 stellten 16% der Krankenhäuser ambulante Operationen zur Verfügung, 1996: Anstieg auf 60%)
- Ausdehnung der ambulanten Langzeitversorgung und Zunahme der häuslichen Betreuung (1990 – 1995: Anstieg der Inanspruchnahme von Heimpflege von 1,9 Millionen auf 3,4 Millionen Patienten)
- Zunahme von Kooperationsverträgen zwischen Krankenhäusern (1985 waren 27% der Gemeindekrankenhäuser Teil des Krankenhaussystems, 1995 waren es mehr als 45%) und Abnahme der stationären Behandlungskosten für Medicare (1980: 67% Aufwendungen für die stationäre Versorgung, 1995: nur noch 49%).

Werden diese Erfahrungen auf Deutschland übertragen, so wäre zu erwarten, dass mit der Einführung eines pauschalierten Entgeltsystems die Leistungstransparenz erhöht und die

Effizienz gesteigert würde. Damit würden nicht nur Kosten eingespart, sondern auch eine leistungsgerechtere Vergütung eingeführt (Garms-Homolová et al., 2002). Der damit geschaffene Anreiz könnte sich aber auch auf die Leistungsqualität stationärer Versorgung auswirken (Laaser et al., 2000; Lauterbach & Lüngen, 2000b). Einige Autoren (Tuschen, 2001; Lauterbach & Lüngen, 2000b) sprechen sich deswegen dafür aus, gleichzeitig mit der Einführung des Fallpauschalensystems *Qualitätssicherungsprogramme* zu implementieren.

Ein weiterer häufig in der Literatur diskutierter Effekt betrifft die *Verlagerung von stationären Leistungen in angrenzende Versorgungsbereiche* (z.B. ambulante oder rehabilitative Behandlung). Ergebnisse einer ersten deutschen retrospektiven Vergleichsstudie legen nahe, dass eine Verkürzung der stationären Liegezeiten tatsächlich mit einem höheren Anteil ambulant behandelter Patienten verbunden sein kann (Laaser et al., 2000).

Für den **psychiatrischen/psychotherapeutischen Versorgungsbereich** liegen inzwischen auch Erfahrungen nach der Einführung von DRG-Systemen vor. Bereits 1993 und 1997 wurden in zwei europäischen Ländern (Ungarn und Österreich) **Fallpauschalensysteme** als Vergütungsmodelle in diesem Bereich implementiert. Diese Systeme orientieren sich im Wesentlichen am allgemeinen DRG-System und enthalten 17 (Österreich) bzw. 16 (Ungarn) diagnosebezogene Fallgruppen (Maylath, 2000; Frick et al., 2001).

In Ungarn werden ca. 50% der psychiatrischen Betten im Akutbereich über das Fallpauschalensystem abgerechnet, in Österreich ca. 70% der staatlich bewilligten Betten (Maylath, 2000; Frick et al., 2001). Die Erfahrungen mit den an den DRG-Systemen orientierten Vergütungsmodellen gehen in den beiden Ländern weit auseinander. Während in Ungarn u.a. eine Reduktion der Verweildauern, eine Verlagerung der Leistungen in angrenzende Bereiche (z.B. Rehabilitationsbetten) bzw. der Ausbau eines eigenen integrierten Versorgungsnetzes der Krankenhäuser mit akuten und chronischen Betten und ein „upcoding“ Verhalten (Zunahme der Vergabe von Diagnosen mit höherem Relativgewicht) zu beobachten waren (Maylath, 2000), wurden in Österreich keine Effekte im Hinblick auf eine Verweildauerreduktion oder auf die Krankenhaushäufigkeit und Wiederaufnahmerate festgestellt (Frick et al., 2001).

Das in Österreich überraschend keine Verkürzung der Verweildauern aufgezeigt werden konnte, hängt vermutlich damit zusammen, dass nur für die mittleren 30% der jeweiligen Verweildauerverteilung eine Pauschale abgerechnet wird und für die verbleibenden 70% aller Aufenthalte eine Steuerung des Entgeltes über die Verweildauer erfolgt (Frick et al., 2001). Die Autoren sprechen in diesem Kontext von einer behutsamen Ausgestaltung des Systems, womit immerhin gelungen ist, ein mit dem somatischen Bereich der stationären Versorgung einheitlich gestaltetes Abrechnungssystem zu implementieren, welches Kostentransparenz ermöglicht.

### 2.7 Überlegungen zu einem Klassifikationssystem in der stationären Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen

Das Gutachten 2003 des Sachverständigenrates für die Konzertierte Aktion im Gesundheitswesen behandelte u.a. auch die Aspekte der Auswirkungen einer DRG-Einführung im Krankenhaus auf die medizinische Rehabilitation (Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen, 2003). In dem Gutachten wurde u.a. geschlussfolgert, dass „durch verkürzte Verweildauern bzw. durch DRG-Effekte im Akutkrankenhaus ‚eingesparte' Krankenhaustage nicht ‚verschwinden', sondern zu einem nicht zu vernachlässigenden Anteil an anderer Stelle, also in den nachgelagerten ärztlich-ambulanten, pflegerischen oder ambulant- bzw. stationär-rehabilitativen Diensten, wieder ‚auftauchen' (zitiert nach Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen, 2003; S. 439). Des Weiteren werden die nachgelagerten ambulanten und rehabilitativen Bereiche durch die frühzeitigeren Krankenhausentlassungen vermutlich mit einer Zunahme der durchschnittlichen medizinischen Fallschwere des Patientenklientels konfrontiert sein. Diese Entwicklungen können aber mit der in der medizinischen Rehabilitation bislang üblichen Vergütungspraxis über tagesgleiche Pflegesätze nicht angemessen begegnet werden.

Deswegen wurde bereits frühzeitig diskutiert, ob nicht auch Rehabilitationsleistungen analog dem DRG-Vergütungssystem abgegolten werden könnten. Die Mehrzahl der Experten war jedoch der Meinung, dass das vorrangig an Diagnosekriterien orientierte DRG-Vergütungssystem das spezifische Leistungsspektrum der medizinischen Rehabilitation konzeptionell nicht angemessen abzubilden vermag und deshalb für diesen Bereich erheblich modifiziert werden müsste. Damit stellt sich die Entwicklungs- und Forschungsaufgabe, ob es auch in dem Bereich der stationären Rehabilitationsbehandlung möglich ist, ein

System zu entwickeln, welches nicht vorrangig an der Diagnose orientiert ist, sondern weitere ressourcenverbrauchsrelevante Kriterien mit einbezieht.

Im folgenden Kapitel soll auf die defizitäre Berücksichtigung von psychischen Störungen bereits etablierter DRG-Systeme und insbesondere des AR-DRG-Systems eingegangen werden. Weiterhin sollen internationale Ansätze zur Entwicklung von Klassifikationssystemen in der Behandlung von Patienten mit psychischen Störungen vorgestellt werden.

# 3 Internationale Ansätze zur Fallgruppenbildung in der Behandlung von Patienten mit psychischen Störungen

Das australische und das adaptierte deutsche DRG-System halten zwei Hauptkategorien für die Verschlüsselung von Erkrankungen aus dem Bereich der psychischen Störungen vor. Zahlreiche empirische Studien und Expertenmeinungen zeigen auf, dass der Bereich der psychischen Störungen in seiner Vielschichtigkeit von den dafür vorgesehenen Kategorien des AR-DRG-Systems nicht erschöpfend abgebildet werden kann. Das führte auf internationaler Ebene zu Bemühungen komplementäre Klassifikationssysteme, die den Anforderungen dieses Gebietes gerecht werden, zu entwickeln. Im Rahmen dieser Bemühungen soll im vorliegenden Abschnitt auch geprüft werden, welche Dimensionen bzw. Kriterien für die Entwicklung eines Fallgruppensystems in der Behandlung psychischer Störungen besonders geeignet sein könnten.

## 3.1 Defizitäre Berücksichtigung psychischer Störungen durch bestehende DRG-Systeme

Zielsetzung zahlreicher, bereits etablierter DRG-Systeme ist es, ein möglichst umfassendes Diagnosen- und Patientenspektrum zu erfassen (Lauterbach & Lüngen, 2000a). Im **AR-DRG-System** werden *psychische Störungen* in der Hauptkategorie 19 kodiert, welche 13 Fallgruppen enthält (Commonwealth Department of Health and Family Services, 1998b) (siehe Tabelle 1). Dazu zählen im Einzelnen Störungen des schizophrenen Formenkreises, Paranoide Störungen, Affektive Störungen, Angststörungen, Ess- und Zwangsstörungen, Persönlichkeitsstörungen und Psychische Störungen des Kindes- und Jugendendalters. Weitere 7 Fallgruppen für die Substanzbezogenen Störungen sind in der Hauptkategorie 20 zusammengefasst (Commonwealth Department of Health and Family Services, 1998b). Mittlerweile liegen erste Versionen des German-DRG-Systems (G-DRG 1.0) vor, die angepasst an das AR-DRG-System zwei Hauptkategorien mit jeweils acht Fallgruppen für die Verschlüsselung von psychischen Störungen vorhalten (MDC 19 „Psychische Störungen“ und MDC 20 „Störungen durch Substanzmissbrauch“).

Eines der Kriterien, welches wichtige Hinweise zur Kostenhomogenität der gebildeten Fallgruppen liefert (Fischer, 1997), ist die durch die einbezogenen Merkmale erzielte **Varianzaufklärung des Ressourcenverbrauchs**. Für die australischen AN-DRGs, Vorläufer des AR-DRG-Systems (vgl. Kapitel 2), werden Varianzreduktionen der Verweildauern von 17% für

die Kategorie MDC 19 berichtet (Fischer, 1997). Das Australian Casemix Clinical Committee (ACCC) publizierte dagegen Daten, wonach die erreichten Varianzreduktionen der beiden psychischen Hauptkategorien (MDC 19 und MDC 20) des australischen Fallgruppensystems bei 11,7% lagen (Buckingham et al., 1998). Im Vergleich dazu werden Varianzaufklärungen der Kosten der *somatischen AR-DRGs* bis 47% erzielt, wenn Patienten mit Extremwerten einbezogen werden und bis 68%, wenn die Berechnungen ohne Extremwerte erfolgen (Fischer, 2002).

Betrachtet man andere international bereits etablierte Fallgruppensysteme, so hält z.B. das *amerikanische DRG-System*, analog zum australischen Klassifikationssystem, zwei Kategorien für Patienten mit psychischen Störungen vor. Zum einen ist das die Kategorie MDC 19 mit neun Fallgruppen (Erkrankungen und Störungen der Psyche) und zum anderen die Kategorie MDC 20 mit sechs Fallgruppen (Alkohol- und substanzbezogene Störungen) (Fischer, 1997). Amerikanische Studien konnten Varianzaufklärungswerte der Verweildauern und der Kosten, gebildet mit den Hauptkategorien psychischer Störungen des amerikanischen DRG-Systems, mit einer Spannweite zwischen 2 bis 15% aufzeigen, wobei die meisten empirischen Befunde in einem Bereich zwischen 3 bis 5% lagen (Taube et al., 1984a, 1984b; English et al., 1986; Schumacher et al., 1986; Horgan & Jencks, 1987; Mitchell et al., 1987; Ashcraft et al., 1989; Horn et al., 1989; Stoskopf & Horn, 1991).

Hauptmerkmal allgemeiner DRG-Systeme ist die Berücksichtigung der **Diagnose** als zentrales Klassifikationskriterium. In Deutschland gibt es mittlerweile einige wenige **empirische Befunde** zum **Zusammenhang der Diagnose von psychisch Erkrankten mit der Behandlungsdauer** in stationärer Behandlung (vgl. auch Kapitel 4). Innerhalb einer Studie zur Basisdokumentation als Grundlage qualitätssichernder Maßnahmen untersuchten von Heymann et al. (2003) Prädiktoren der stationären Behandlungsdauer. Die Autoren bezogen hierbei 17 Einrichtungen aus unterschiedlichen stationären Versorgungsbereichen von Patienten mit psychischen/psychosomatischen Störungen mit einer Gesamtstichprobe von N = 17.860 Patienten in die Studie ein. Es konnten verschiedene psychosoziale, klinische und diagnostische Variablen identifiziert werden, die einen signifikanten Zusammenhang mit der Behandlungsdauer aufwiesen (vgl. Kap. 4). Deswegen schlussfolgerten die Autoren, dass eine *diagnosebezogene Verweildauer*, so wie das AR-DRG-System impliziert, nur als ein Wirkfaktor neben anderen, den Ressourcenverbrauch mitbestimmenden Variablen zu verstehen ist. Auch Jagdfeld et al. (2002) konnten anhand eines eher geringen Stichprobenumfanges von N = 61 stati-

onären Patienten aufzeigen, dass z.B. Komplikationen und Komorbiditäten von Patienten mit psychischen/psychosomatischen Störungen bei der Kodierung mittels des AR-DRG-Systems unberücksichtigt bleiben. Mestel et al. (2002) zeigten an einer Gesamtstichprobe von 1.953 Patienten aus einer Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen, dass weitere Variablen, die mit der Basisdokumentation erhoben wurden, einen signifikanten Zusammenhang zur Behandlungsdauer aufwiesen. Dagegen konnte diese Assoziation nur für wenige Diagnosen gefunden werden. Schließlich untersuchte Richter (Richter, 1999) in einem Krankenhausbetriebsvergleich für Kliniken aus dem Bereich der Psychiatrie und Psychotherapie den Zusammenhang zwischen Diagnose und Alter mit der Behandlungsdauer der Patienten. Er kam zu dem Ergebnis, dass die beiden Variablen 10,5% der Varianz der Behandlungsdauer aufklärten und weitere routinemäßig erhobene soziodemographische und klinische Merkmale nur zu einer marginalen Erhöhung der Varianzaufklärung beitrugen (um 0,8%). Dieses Ergebnis deutet darauf hin, dass weitere Kriterien (außer der Diagnose und des Alters der Patienten) den Ressourcenverbrauch in stationärer Behandlung substantiell mitbestimmen.

Nicht nur das der Zusammenhang zwischen der psychischen Diagnose eines Patienten und der Behandlungsdauer eher gering ist, auch **weitere Probleme**, die sich auf dieses Merkmal beziehen, könnten mit einer Einführung des AR-DRG-Systems für den Bereich psychischer Störungen verbunden sein. Elsner et al. (2003) verdeutlichten dies an Beispielen aus der Krankenhausbehandlung von Patienten mit Substanzmissbrauch. Sie stellten fest, dass die Verschlüsselung der Diagnosen, wenn sie analog zum australischen Fallgruppensystem vorgenommen würde, für diesen Bereich defizitär ist: Im AR-DRG-System werden die ersten 3 Stellen der ICD-10 Kodierung übernommen. Das würde einerseits bedeuten, dass das gesamte Störungsspektrum für diesen Bereich nicht abgebildet wird (z.B. F 10.00 Intoxikation mit Alkohol ohne medizinische Komplikationen, F 10.03 Delir im intoxikierten Zustand) und andererseits, dass keine angemessene Kostengewichtung vorgenommen werden kann, da relevante Nebendiagnosen nicht mit verschlüsselt werden. Aus Sicht der Autoren würde sich damit eine erhebliche wirtschaftliche Schieflage ergeben, die mit einer Gefährdung der geeigneten Behandlungsmaßnahmen einhergehen könnte.

Aufgrund dieser empirischen Befunde und nach Expertenmeinungen (Gerdelmann, 1997; Lauterbach & Lüngen, 2000a; Kruckenberg et al., 2001) zeigt sich demnach, dass eine **Fallgruppierung** von Patienten mit psychischen Störungen **primär nach Diagnose**, wie im AR-

DRG-System vorgenommen, zu keiner adäquaten Abbildung des Ressourcenverbrauches führt und die Varianz der Behandlungskosten nicht ausreichend erklären kann. Dazu trägt der Umstand bei, dass Therapieverläufe von Patienten mit psychischen Störungen trotz gleicher Haupt- und Nebendiagnose sehr unterschiedlich (Gerdelmann, 1997) und von zahlreichen weiteren Einflussfaktoren (z.B. Akzeptanz der Therapieangebote durch die Patienten, Therapiemotivation der Patienten) abhängig sein können (Gerdelmann, 1997; Mestel et al., 2002). Daher wurde zunächst der Bereich der stationären Versorgung von Patienten mit psychischen Störungen von der Einführung des AR-DRG-Systems ausgenommen (Bundesministerium für Gesundheit, 2002). Das gilt jedoch nicht für die Vergütung von Leistungen psychisch Kranker, die sich z.B. in internistischer stationärer Krankenhausbehandlung befinden (Häuser et al., 2004).

Die **vertretenden Fachgesellschaften** nehmen dazu unterschiedliche Positionen ein. Während die Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) und die Arbeitsgemeinschaft für Neuropsychopharmakologie und Pharmakopsychiatrie (AGNP) den Ausschluss aus dem DRG-System begrüßen (Fritze et al., 1999), spricht sich die Deutsche Gesellschaft für Psychotherapeutische Medizin (DGPM) für eine Teilnahme aus (Deutsche Gesellschaft für Psychotherapeutische Medizin (DGPM), 2001). Aufgrund dieser differenten Positionen der Fachgesellschaften, wegen der momentan noch vorläufigen Ausnahmeregelung und der nicht absehbaren mittel- bzw. langfristigen Perspektive auch für die stationäre Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen (vgl. hierzu Abschnitt 2.7) bedarf es der Auseinandersetzung mit DRG-Systemen für die stationäre Behandlung von Patienten mit psychischen Störungen, wie bereits Burgmer und Freyberger (2002) konstantierten. In diesem Kontext sollte aber zunächst überprüft werden, ob eine Fallgruppierung in diesem Indikationsbereich möglich ist und der Frage nachgegangen werden, welche Kriterien substantiell den Ressourcenverbrauch psychisch Erkrankter in stationärer Behandlung mitbestimmen.

### 3.2 Internationale Klassifikationssysteme für Patienten mit psychischen Störungen

Die in vielen empirischen Studien aufgeführten **Varianzaufklärungswerte** der Verweildauer/Behandlungskosten der allgemeinen DRG-Systeme in einem Bereich von 2 bis 15% lassen darauf schließen, dass die in den bereits etablierten DRG-Systemen vorhandenen Variablen

den Ressourcenverbrauch für Patienten mit psychischen Störungen nicht ausreichend abbilden können und dass weitere Merkmale den Ressourcenverbrauch substantiell mitbestimmen. Aus diesem Grund ist es notwendig, ein adäquates, über den Ansatz allgemeiner Klassifikationssysteme hinausgehendes Fallgruppensystem zu entwickeln. Wegen der mangelnden Differenzierungsfähigkeit der bisherigen *allgemeinen* DRG-Systeme bezüglich psychischer Erkrankungen (Horn et al., 1989; Buckingham et al., 1998) gilt es zu prüfen, inwieweit Bemühungen innerhalb der Psychiatrie zu geeigneteren Klassifikationssystemen geführt haben, die in der Lage sind, Patienten mit vergleichbarem Ressourcenverbrauch zusammenzufassen. Dabei soll insbesondere überprüft werden, welche **Klassifikationskriterien** bei der Bildung von Fallgruppen innerhalb dieser Systeme einen hohen Stellenwert einnehmen, um sie für die Entwicklung eines solchen Ansatzes nutzbar zu machen.

Nachfolgend werden **internationale Klassifikationssysteme** aus dem Bereich der *psychiatrischen/psychotherapeutischen* Behandlung beschrieben und anschließend hinsichtlich ihrer Eignung zur Übernahme von Klassifikationskriterien für ein nationales Patientenklassifikationssystem bewertet.

Tabelle 2 gibt einen Überblick über die hier berücksichtigten international entwickelten Klassifikationssysteme im Bereich der Erwachsenen-Psychiatrie, welche im Folgenden näher vorgestellt werden sollen. Die Kinder- und Jugendpsychiatrie wurde ausgeklammert, da Erfahrungen aus diesem Bereich für die Entwicklung eines Klassifikationssystems für Erwachsene nicht ausreichend transferierbar erscheinen.

Tabelle 2: Übersicht über internationale Klassifikationssysteme in der Psychiatrie

| *Bezeichnung* | *Autoren* | *Anwendungsbereich* | *Effektive Stichproben-größe* | *Anzahl der Fall-gruppen* | *Varianzauf-klärung (in Prozent)*[1] |
|---|---|---|---|---|---|
| **M**ental **H**ealth **C**lassification and **S**ervice **C**osts **P**roject | Buckingham et al. (1998) | stationäre und ambulante Versorgung (Australien) | 3.426 (stationär) | 8 (stationär) | 16,3[2] |
| **N**ew **Z**ealand **M**ental **H**ealth **C**lassification **a**nd **O**utcomes **S**tudy **(NZ-CAOS)** | Gaines et al. (2003) | stationäre und ambulante Versorgung (Neuseeland) | 1.577 (stationär) | 21 (stationär) | 15,2[2] |
| **P**sychiatric **S**everity of **I**llness | Horn et al. (1989) | stationärer psychiatrischer Bereich (USA) | 1.418 | 4 | 34 – 50[1] |
| **P**sychiatric **P**atient **C**lassifica-tion **S**ystem | Ashcraft et al. (1989) | stationärer psychiatrischer Bereich und substanzbezogene Störungen (USA) | 105.727 | 74 | 18[1] |
| **S**ystem for **C**lassification of **I**n-Patient **P**sychiatry | Hirdes et al. (2003) | stationärer psychiatrisch/psycho-therapeutischer Bereich | 1.998 | 47 | 33 – 38[2] |
| **P**atient **M**anagement **C**atego-ries | Young et al. (1985) | Akutversorgung inklusive stationärer psychiatrischer Bereich (USA) | keine Angabe | 47 Module mit 830 Patienten-kategorien | 18[1] |
| **A**lternate **DRGs** | Taube et al. (1984b) | Stationärer psychiat-risch/psychotherapeutischer Bereich | 12.618 | 22 | 11,8[1] |
| **C**linically **R**elated **G**roups und **S**ystemetrics´s **D**isease **S**taging | Mitchell et al. (1987) | Stationärer psychiat-risch/psychotherapeutischer Bereich | 22.000 | 16 | 9,3[1,2] (CRG)<br>9,2[1,2] (SDS) |

**Anmerkungen:** Tabelle in Anlehnung an (Fischer, 1997). Fett markierte Buchstaben ergeben die jeweiligen Abkürzungen der Klassifikationssysteme. [1] Varianzaufklärung der Aufenthaltsdauer, [2] Varianz-aufklärung der Kosten

### *3.2.1 Mental Health Classification and Service Cost Project (MH-CASC)*

Buckingham et al. (1998) setzten sich mit dem Mental Health Classification and Service Costs Project (MH-CASC) in Australien zum Ziel, ein Klassifikationssystem für alle Versorgungsbereiche von Patienten mit psychischen Störungen (stationäre, teilstationäre und ambulante Therapie) zu entwickeln. Im Folgenden werden nur Ergebnisse berichtet, die sich auf die Gruppe der stationär behandelten Patienten beziehen, da diese mit der hier zu untersuchenden Stichprobe den größten Überschneidungsbereich aufweist.

Die Autoren bezogen 22 Einrichtungen mit einer Gesamtstichprobengröße von ca. 18.000 Patienten in die Entwicklung des Klassifikationssystems ein. Zunächst wurden mithilfe einer

umfangreichen Literaturrecherche und aufgrund von Empfehlungen eines Expertenpanels Patientenmerkmale identifiziert, die möglicherweise den Ressourcenverbrauch determinieren könnten. Die Erhebung der Variablen (z.B. psychische Diagnosen[3], Schweregrad der Erkrankung und Funktionsniveau des Patienten anhand von Fremdeinschätzungen) erfolgte durch das Personal der Einrichtung am Anfang der Behandlung, weiterhin kontinuierlich im 14-tägigen Turnus sowie am Ende der Behandlung. Zur Einschätzung des Ressourcenverbrauches dokumentierte das Personal außerdem täglich angewandte patientenbezogene, nicht-patientenbezogene (z.B. Forschungsaktivitäten) und generelle Leistungen (Restkategorie).

Abbildung 4 zeigt den Fallgruppenbaum[4], den Buckingham et al. (1998) für die Patienten, die ihre stationäre psychiatrische/psychotherapeutische Behandlung beendet hatten, extrahierten. Es wurden 8 Fallgruppen identifiziert, welche insgesamt 16,3% der Varianz des Ressourcenverbrauches aufklärten. Die Fallgruppen wurden aufgrund der Kriterien *Alter*, *Diagnose*, *Symptombelastung*, *Status der Einweisung* und *Activity of Daily Living - Index (ADL-Index)* (für Patienten über 65 Jahre) gebildet (Abbildung 4). Wie aus Abbildung 4 ersichtlich ist ein höherer Ressourcenverbrauch mit einem höheren Schweregrad (gemessen mit der Health of the Nation Outcome Scales, HoNOS, den Resource Utilization Groups, RUG und dem Activity of Daily Living-Index, ADL und unfreiwilliger Aufnahme in die psychiatrische Einrichtung verbunden. Ein geringerer Ressourcenverbrauch ist dagegen mit leicht ausgeprägtem Schweregrad assoziiert. Die Patientengruppe, welche am meisten Ressourcen verbrauchte, ist dadurch gekennzeichnet, dass sie im Durchschnitt über 85 Jahre alt waren. Die kostengünstigste Gruppe war durch Patienten charakterisiert, die durchschnittlich unter 65 Jahre alt waren und keine Diagnosen aus dem Bereich der Schizophrenie, Depressiven Störung oder Essstörung aufwiesen (siehe Abbildung 4).

---

[3] Psychiatrische Diagnosen auf der Grundlage von 16 Hauptkategorien und 61 individuellen Kategorien, die auf dem ICD-10 basieren.

[4] Eine Erläuterung zur Konstruktion der Fallbäume wird in Kap. 9, Abschnitt 9.6.2.1 dieser Arbeit gegeben.

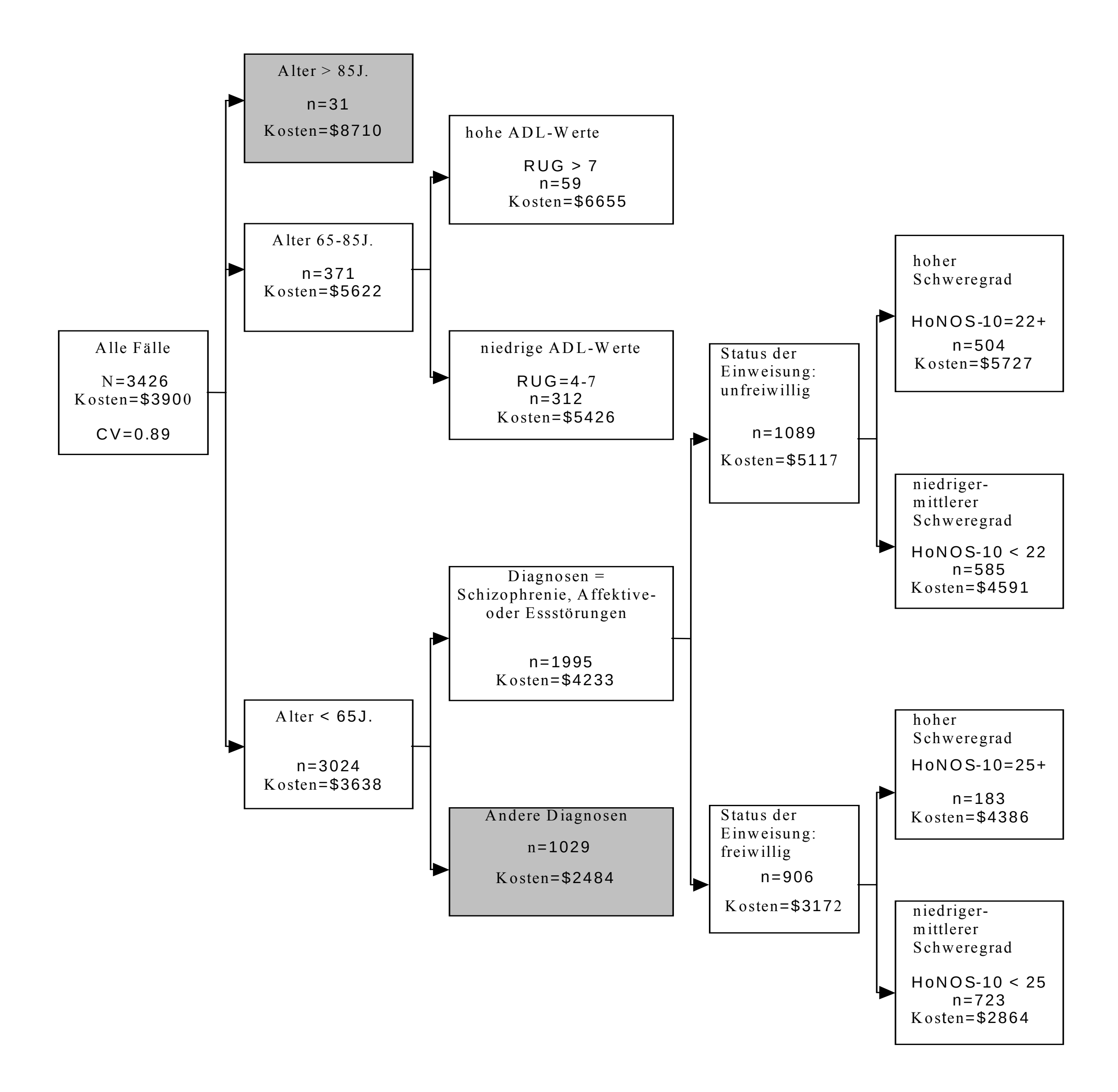



Abbildung 4: Fallgruppenbaum des australischen Klassifikationssystems (MH-CASC, Buckingham et al., 1998) für die Patienten mit einer bereits abgeschlossenen stationären Behandlung; Anmerkungen: HoNOS = Health of the Nation Outcome Scales, N = Gesamtstichprobengröße, n = Stichprobengröße, durchschnittliche Kosten in Dollar, CV = Varianzkoeffizient (coefficient of variance), RUG = Resource Utilization Groups, ADL = activity of daily living Index; grau hinterlegt = ressourcenintensivste und ressourcengünstigste Gruppe

Im damals verwendeten, allgemeinen AR-DRG-System wurden für den Bereich der stationären psychiatrischen Versorgung acht Fallgruppen bei einer Stichprobengröße von 2.541 Patienten identifiziert. Das MH-CASC Projekt klassifizierte elf Fallgruppen für den stationären psychiatrischen Bereich (für Therapie beendet: Erwachsene und Kinder/Jugendliche) auf der Basis einer Stichprobengröße von 3.758 Patienten. Das Klassifikationssystem des MH-CASC Projektes erklärte den Ressourcenverbrauch für diese elf Fallgruppen effektiver als das AR-DRG-System (17,3% vs. 9,9%).

### 3.2.2 New Zealand Mental Health Classification and Outcomes Study (NZ-CAOS)

Beauftragt durch das Health Research Council und maßgeblich beeinflusst durch die Ergebnisse des australischen MH-CASC Projektes, wurde 1999 ein Projekt in Neuseeland mit der Zielsetzung initiiert, sowohl ein Fallgruppensystem für den Bereich von Patienten mit psychischen Störungen zu entwickeln als auch Erfahrungen im Bereich der Routineerhebung mit Outcomeinstrumenten zur Verfügung zu stellen (Gaines et al., 2003).

Auf der Basis von fünf Kriterien (z.B. Repräsentativität oder flächendeckende Versorgung) wurden acht Distrikte für die Studie ausgewählt, die insgesamt 22% der Behandlung psychisch Kranker in Neuseeland abdeckten. An der Erhebungsphase nahmen 103 Teams über einen Zeitraum von 6 Monaten teil. Wie im australischen MH-CASC Projekt bildeten drei Datensäulen die Grundlage der Erhebung: Patientencharakteristika, direkte Kosten der Einrichtungen (z.B. Therapieleistungen) und indirekte Kosten (z.B. Lehre, Fortbildungen). Bei der Untersuchung zu Beginn und am Ende der Behandlung wurden größtenteils die bereits bei der MH-CASC Studie eingesetzten Instrumente verwendet (u.a. auch der HoNOS).

Bezüglich der Ergebnisdarstellung sollen auch hier, analog zum vorigen Abschnitt, nur die Fallgruppen der stationären Patienten mit bereits abgeschlossener Behandlung berichtet werden. Zunächst wurde das Klassifikationsmodell des MH-CASC Projektes überprüft, welches sich aber aufgrund unterschiedlicher Proportionen ethnischer Gruppen als unzureichend für neuseeländische Bedürfnisse erwies. Abbildung 5 vermittelt einen Überblick über die in Neuseeland gebildeten 12 Fallgruppen, wobei sich 6 der 12 Fallgruppen hinsichtlich dreier ethnischer Gruppen (Maori, Pacific Island und Europäer/andere) noch weiter auffächern. Die 12 Fallgruppen klärten insgesamt 14,74% (ungetrimmt) bzw. 15,23% (getrimmt) der Varianz der Behandlungskosten auf. Einer der bedeutsamsten Prädiktoren des Ressourcenverbrauchs war das Item 6 der HoNOS („Halluzinationen und Wahnvorstellungen“), gefolgt vom Alter der Patienten, dem Behandlungsfokus („Focus of Care“), dem Einweisungsstatus, der ethnischen Zugehörigkeit und der HoNOS Subskala „Impairment“ (Item 4 und 5). Kritisch muss allerdings angemerkt werden, dass einige gebildete Fallgruppen keine Fallzahlen (z.B. n = 0) oder nur sehr geringe Fallzahlen (z.B. n = 10) aufwiesen, so dass die hier berichtete Varianzaufklärung sicherlich überschätzt ist.

Auch in **Großbritannien** wurde ein Projekt (**Mental Health Groupings, MHG**) mit dem Ziel initiiert, ein Gruppierungssystem unter Berücksichtigung von Patientenbedürfnissen, Interventionen und dem Outcome von Patienten mit psychischen Störungen zu entwickeln (Carthew et al., 2003). Da dieses Projekt und seine Ergebnisse nicht zentral für die vorliegende Arbeit sind, wird im Folgenden auf eine ausführliche Darstellung verzichtet. Es soll allerdings angemerkt werden, dass auch in diesem Projekt, genauso wie im australischen MH-CASC und im neuseeländischen NZ-CAOS, u.a. der HoNOS an einer Gesamtstichprobe von 6000 Patienten in 6 Bezirken Großbritanniens in einem Zeitraum von einem halben Jahr erhoben wurde, welche zur Bildung von so genannten „Health Benefit Groups“[5] beitrug. Konkret bedeutete das, dass die Diagnosegruppen des ICD-10 (z.B. F00 Demenz) in drei Schweregrade im HoNOS unterteilt wurden: Hoher Ausprägungsgrad im HoNOS (Gesamtscore > 20), moderater Schweregrad (Gesamtscore 11-20) und niedriger Schweregrad (Gesamtscore < 10). Das führte zu einer Bildung von 69 Gruppen. Bei der Fallgruppenbildung spielte im Gegensatz zu anderen Klassifikationssystemen die Kategorisierung von Interventionen (z.B. Krisen- oder Kurzzeittherapie) und die aufgewendeten Ressourcen (z.B. Ganztags- oder Halbtagsbetreuung) eine zentrale Rolle. Das führte insgesamt zu einer Fallgruppenbildung von 130 Gruppen.

[5] „Health Benefit Groups“ unterteilen Patienten bezüglich ihres Versorgungsbedarfs und dem erwarteten Outcome (wenn ähnliche Interventionen gegeben wurden).

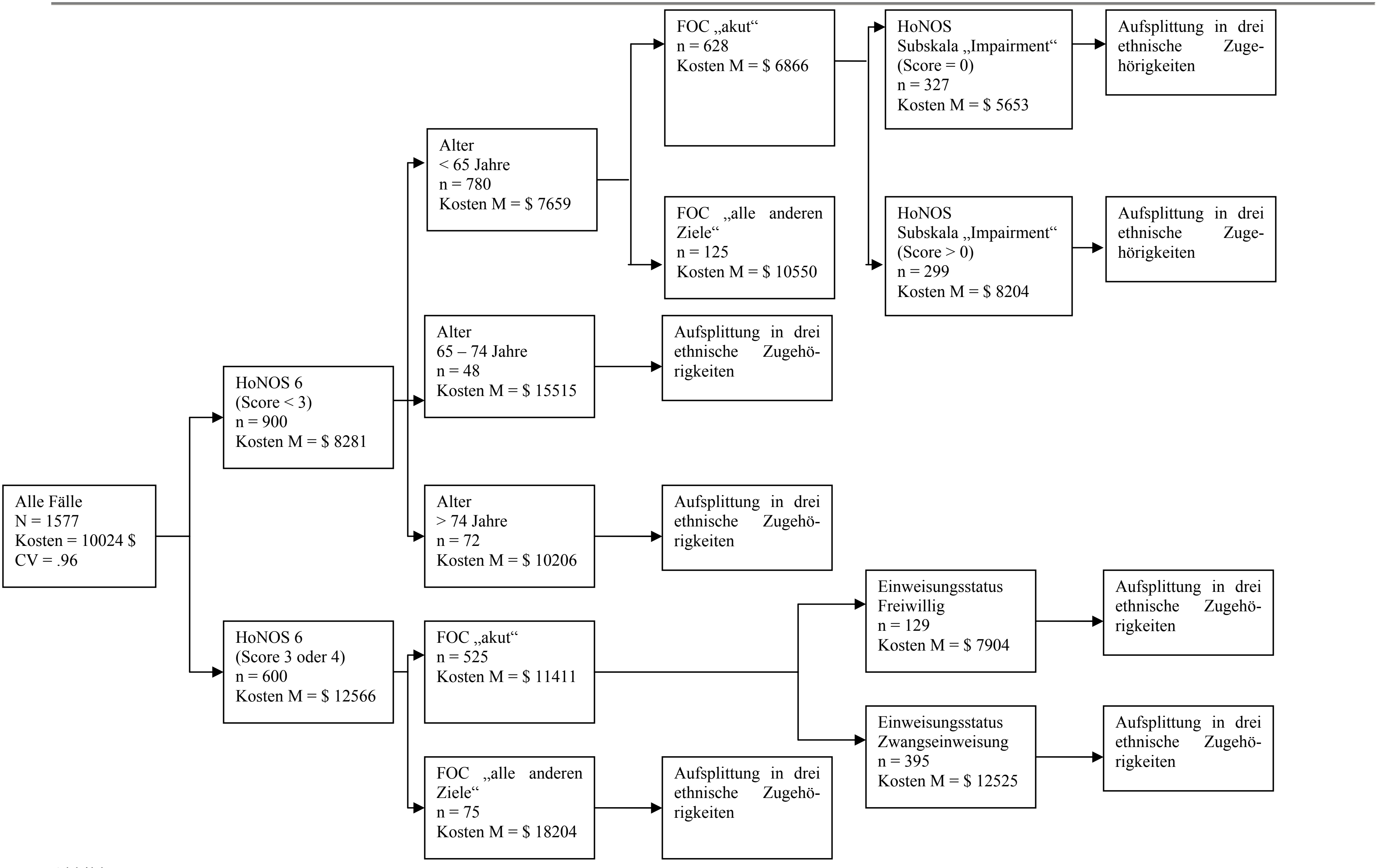



Abbildung 5: Fallgruppenbaum des neuseeländischen Klassifikationssystems (NZ-CAOS, Gaines et al., 2003). FOC = Focus of Care; HoNOS = Health of the Nation Outcome Scales

### *3.2.3 Psychiatric Severity of Illness Index (PSI)*

Ziel der Studie von Horn et al. (1989) in den **USA** war die Entwicklung einer Klassifizierung von stationären psychiatrischen Patienten zu homogenen Fallgruppen aufgrund der **Schwere ihrer Erkrankung**. In der Studie wurde untersucht, ob eine Bildung von Fallgruppen basierend auf dem Psychiatric Severity of Illness Index (PSI) zu einer homogeneren Abbildung der Verweildauern führte als die Fallgruppenbildung des allgemeinen amerikanischen DRG-Systems.

In die Studie von Horn et al. (1989) wurden zehn Krankenhäuser (akute psychiatrische Stationen) mit 1.418 Patienten einbezogen. Davon wurden 1.173 Patienten einer der neun Fallgruppen des amerikanischen DRG-Systems MDC 19 zugeordnet. Als Gruppierungskriterium wurde der PSI herangezogen, der das Ausmaß der Schwere der Erkrankung auf der Basis von vier Kategorien (leicht belastet bis schwer belastet) unter Verwendung von sieben Dimensionen erfasst: *Erkrankungsstadium*, *Komplikationen während des Aufenthaltes*, *Interaktionsverhalten*, *Pflegeaufwand*, *Soziale Unterstützung*, *Therapiecompliance* und *Behebung der akuten Symptomatik*. Die sieben Dimensionen sind in Tabelle 3 näher beschrieben und erläutert. Der Gesamtscore und damit die Zuordnung zu einer PSI Fallgruppe wurde durch das Ausmaß der Schwere über alle sieben Dimensionen ermittelt.

Die Varianzreduktion, die auf der Basis des PSI-Systems erzielt wurde, war gegenüber dem amerikanischen DRG-System beachtlich. Horn et al. (1989) zeigten, dass vier PSI-Fallgruppen 42% der Varianz in der Verweildauer erklärten, wenn sie zur Verbesserung der amerikanischen DRGs verwendet wurden (PSI kombiniert mit DRGs). Die neun Fallgruppen der MDC 19 Kategorie des amerikanischen DRG-Systems klären allein dagegen nur 7% der Varianz der Aufenthaltsdauer in der Studie von Horn et al. (1989) auf. Allerdings sollte bei diesem Ergebnis berücksichtigt werden, dass es sich nicht um die Konstruktion eines prospektiven Systems handelt, sondern bei Horn et al. (1989) wurden auch Komplikationen während des stationären Aufenthaltes und der pflegerische Aufwand in dem Schweregradindex mitberücksichtigt. Deswegen kann vermutet werden, dass die beachtliche Höhe der Varianzaufklärung auch durch die Erhebung von therapieverlaufsbezogenen Kriterien erklärlich wird.

Einige Jahre später entwickelten Stoskopf und Horn (1991) eine **Computerversion des Psychiatric Severity Index** (CPSI), um Schwachstellen der manuellen Version auszuräumen

(z.B. Manipulierbarkeit der Daten). Die Utilität dieses Konzeptes wurde dann in einer Studie mit Patienten der Diagnosegruppen Schizophrenie und Affektive Störungen (ohne Komorbidität) untersucht (Stoskopf & Horn, 1992).

Tabelle 3: Beschreibung der Dimensionen des PSI-Systems (Horn et al., 1989)

| Dimensionen des PSI-System | Beschreibung der Dimensionen |
|---|---|
| 1. Erkrankungsstadium | ▪ Bezieht sich auf das Ausmaß des zugrundeliegenden Problems, das sich in Symptomen und Beeinträchtigung manifestiert. |
| 2. Komplikationen während des Aufenthalts | ▪ Komplikationen werden als unerwünschte Ereignisse während eines stationären Aufenthaltes verstanden, die negative Auswirkungen auf die Gesundheit des Patienten haben. Diese Ereignisse schließen Vorfälle ein, die mit der Therapie einhergehen (z.B. Drogenintoxikation). |
| 3. Komorbiditäten | ▪ Komorbiditäten können auf zwei verschiedene Art und Weisen auftreten: a) vorausbestehende Bedingungen oder Probleme, die keine klare Beziehung zur psychischen Hauptdiagnose haben (z.B. Magengeschwür bei einem schizophrenen Patienten) oder b) vorausbestehende Bedingungen oder Probleme, die in irgendeiner Art mit einer hohen Wahrscheinlichkeit ursächlich oder bezogen auf die Hauptdiagnose vorkommen können (z.B. histrionische Persönlichkeitsstruktur eines Patienten mit Anpassungsstörung). |
| 4. Pflegeaufwand | ▪ Beschreibt den Grad der pflegerischen Leistungen während der stationären Behandlung. |
| 5. Soziale Unterstützung | ▪ Beschreibt in wie weit andere, für den Patienten bedeutsame Personen dem Patienten emotionale Unterstützung geben, empfänglich für die Wünsche des Patienten sind und auf den Patienten eingehen. |
| 6. Therapiecompliance | ▪ Das Ausmaß der Reaktion auf die Therapie bezieht sich darauf, wie der Patient auf die Therapie anspricht und auf die Effekte, die Komplikationen und schwierige Interaktionen mit sich bringen könnten. |
| 7. Behebung der akuten Symptomatik | ▪ Bezieht sich auf das Ausmaß der Verbesserung der akuten Symptomatik der Hauptdiagnose oder der Komplikationen. |

Die CPSI-Kriterien für die Diagnosen Schizophrenie und Affektive Störungen wurden von klinischen Experten erarbeitet (Diplom-Psychologen und Fachärzten für Psychiatrie) und umfassten vier verschiedene Bereiche: 1) *psychischer Status*, 2) *Erkrankungsvorgeschichte*, 3) *medizinische Komplikationen* und 4) *psychosoziale Faktoren*. Dieser aufgestellte Kriteriumskatalog wurde anschließend erneut von mehreren klinischen Experten geprüft. Tabelle 4 enthält ein Beispiel für die Variablen, die für den Bereich der affektiven Störungen „psychischer Status“ aufgestellt wurden.

Tabelle 4: Beispielkategorie „psychischer Status“ des CPSI-Systems mit Beschreibung der Kriterien für die Affektive Störung (Stoskopf & Horn, 1992)

| **Ausprägungsgrade der Symptomschwere** | | | |
|---|---|---|---|
| **1 (normale bis milde Symptomatik)** | **2 (mittelmäßig ausgeprägte Symptomatik)** | **3 (stark ausgeprägte Symptomatik)** | **4 (lebensbedrohliche Symptomatik)** |
| | Halluzinationen/Wahnvorstellungen mit bedeutsamen Verhaltenskonsequenzen | Verhaltensmanifestationen der Halluzinationen/Wahnvorstellungen | drohende Gefahr bestehende Halluzinationen/Wahnvorstellungen auszuagieren und damit andere oder sich selbst zu gefährden; Halluzinationen/Wahnvorstellungen behindern die Behandlung potentiell lebensbedrohlicher Erkrankung |
| Psychomotorische Retardierung | mangelhafte Selbstpflege mit Rückzugsverhalten /Disorganisation | Stark ausgeprägtes Rückzugsverhalten mit Ernährungsmangel | |
| | in Abständen auftretende Suizidgedanken | persistierende Suizidgedanken | aktive Suizidalität |
| | aggressives/feindseliges/dysphorisches/gereiztes Verhalten | kampflustig | Fremdgefährdung |
| milde Veränderung in der Stimmungslage, Lebenssinn vorhanden | impulsiv, aufgeregt; eher schlechter Stimmung | schwerer ausgeprägte Agitation; Ablenkbarkeit; ohne Beherrschung | manisches Delirium |

Im Anschluss wurde ein Verfahren zur Erhebung des Gesamtschweregrades entwickelt, wobei die Ziffer, welche die Messung des Schweregrades repräsentierte, an die ICD-9 Diagnose angefügt wurde. An der Studie war ein großes, städtisches Krankenhaus beteiligt. Die Stichprobengröße betrug insgesamt N = 304 Patienten (n = 61 Patienten mit einer schizophrenen Störung, n = 243 Patienten mit einer affektiven Störung). Die CPSI-Kriterien wurden sowohl zu Beginn der Behandlung (nach 48 Stunden), auf einem Teil der Informationen basierend, als auch nach fortgeschrittener Behandlung, auf der medizinischen Akte basierend (maximale Schweregraderfassung), eingeschätzt. Zur Erfassung der Güte der Schweregradmessung wurden Interraterreliabilitäten zwischen zwei Fachärzten für Psychiatrie, einem Diplom-Psychologen und der Studienraterin bestimmt. Zusätzlich wurden 50 Prädiktoren erhoben (u.a. Alter, Geschlecht, Obdachlosigkeit, frühere Krankenhausbehandlungen, Alter bei der ersten Krankenhausbehandlung, Anwendung von Elektrokrampftherapie), von denen 22 Prädiktoren in die Analysen eingingen.

Die Interraterreliabilitäten der CPSI-Kriterien können für beide Patientengruppen als gut bezeichnet werden (r = .82 - .86) (Stoskopf & Horn, 1991). Innerhalb der amerikanischen DRG Fallgruppe für „Psychotische Erkrankungen (DRG 430)" konnten neun Prädiktoren (CPSI-Score, Kostenträgerstatus, Gabe von Tranquilizern bei Entlassung, Entlassungsfähigkeit, Einnahme von Antidepressiva zu Beginn der Behandlung und Alter von 25-64 Jahren) insgesamt 32,5% der Varianz der Behandlungsdauer erklären (Stoskopf & Horn, 1992). Regressionsanalysen für die beiden Diagnoseuntergruppen zeigten Varianzaufklärungsraten für die affektiven Störungen von 27% und für die Patienten mit schizophrener Störung von 70%. Kritisch anzumerken ist, dass bei der Interpretation dieses Ergebnisses das Verhältnis der eingegangenen Prädiktoren zur Stichprobengröße beachtet werden sollte. Weiterhin sollte dabei berücksichtigt werden, dass mit der Schweregradeinteilung bereits eine Diagnosezuordnung verbunden ist und dieses Merkmal anscheinend zweifach in die Analysen eingeht (über die Fallgruppe und über die Schweregradeinteilung).

### *3.2.4 Psychiatric Patient Classification System (PPC)*

Ashcraft et al. (1989) strebten an, ein psychiatrisches Klassifikationssystem in den USA zu entwickeln, dessen klinische Aussagekraft höher sein und das in ihrem Ressourcenverbrauch homogenere Fallgruppen abbilden sollte als das vorhandene amerikanische DRG-System.

Zunächst wurde ein Expertenpanel gebildet, welches aus Psychiatern und Psychologen mit Lehrtätigkeit bestand. Aufgabe der Panelmitglieder war die Identifizierung von klinisch relevanten Kriterien aufgrund ihrer Erfahrung und ihrer Kenntnisse über den Literaturstand. Die Kriterien wurden letztendlich über eine modifizierte Delphi-Methode zusammengetragen. Ashcraft et al. (1989) entwickelten einen kurzen Fremdeinschätzungsfragebogen, der alle gewonnenen Kriterien aus der Delphi-Runde enthielt. Diese bezogen sich auf drei Dimensionen: Verhaltensfaktoren, Funktionsfähigkeit und soziale Charakteristika. Zur Konstruktion des Systems erfolgte die Datenerhebung über einen Zeitraum von neun Monaten, wobei auch die Entlassungsdaten zur Analyse herangezogen wurden. Die Daten des Fragebogens wurden mit der Hauptdiagnose, mit den Prozeduren und demographischen Angaben verknüpft. Die Gesamtstichprobengröße betrug 116.191 Patienten. Von der Auswertung wurden diejenigen Patienten ausgeschlossen, die entweder nur 1 Behandlungstag oder mehr als 100 Tage auf der Station verweilten (n = 5.324) und diejenigen, die irregulär (Entlassung gegen ärztlichen Rat, Todesfall usw.) entlassen wurden (n = 5.140).

Im ersten Schritt wurden die Behandlungsfälle einer von 12 Hauptdiagnosegruppen zugeordnet. Danach erfolgte innerhalb der Gruppen eine weitere Partitionierung anhand folgender Kriterien: *Anzahl der somatischen Hauptdiagnosen*, *psychische Komorbidität*, *Alter*, *Verwirrtheitszustand*, *Zuweisung zu einer für Posttraumatische Belastungsstörung spezialisierten Station*, *Ersteinweisung*, *ADL-Index*, *Schwere der Erkrankung bei Aufnahme* und *Aufnahme zur Entgiftung* mit einer Behandlungsdauer von weniger als sieben Tagen. Von den aufgeführten Variablen wurden in 98% der Fälle *nur zwei Kriterien* zur Klassifikation innerhalb der Hauptdiagnosegruppe benutzt (die Anzahl der somatischen Hauptdiagnosen und das Ausmaß der Schwere der Erkrankung bei Aufnahme).

Innerhalb der 12 Hauptdiagnosegruppen wurden insgesamt 74 Fallgruppen identifiziert. Im Gegensatz zu den amerikanischen DRGs konnte allein für die Diagnosegruppe „Substanzbezogene Störung“ des PPC-Systems die Varianzaufklärung von 1,6% auf 32,0% erhöht werden. Über alle 12 Diagnosegruppen des PPC-Systems mit insgesamt 74 Fallgruppen wurde die Varianz der Verweildauer der Patienten zu 18% erklärt (DRGs: 3,3% in der Studie von Ashcraft et al., 1989).

#### *3.2.5 System for Classification of In-Patient Psychiatry (SCIPP)*

Zielsetzung des Projektes von Hirdes et al. (2003) war die Entwicklung eines **Fallgruppensystems zur Erfassung des täglichen Ressourcenverbrauches** in allen stationären psychiatrischen Einrichtungen in Ontario (**Kanada**). Darunter wurden 34 Einrichtungen dreier kanadischer Provinzen mit kurzen und langen Behandlungsdauern sowie forensische und geriatrische Abteilungen subsumiert.

Ingesamt wurde eine Stichprobengröße von 1.998 Patienten untersucht, wobei 38% der Patienten eine schizophrene oder andere psychotische Störung als Erstdiagnose aufwiesen, gefolgt von Patienten mit depressiven Störungen (36%). 47% der untersuchten Patienten befand sich in stationärer psychiatrischer Behandlung. Alle Mitarbeiter der Station kodierten die erbrachten Leistungen in drei verschiedene Kategorien: 1) patientenspezifische Zeit (z.B. Beratungsgespräche), 2) nicht-patientenspezifische Zeit (z.B. Kontrollgang durch die Station und Kontrolle der Überwachungsmonitore) und 3) nicht-patientenspezifische Zeit, die mit nichtklinischen Tätigkeiten verbracht wurde (z.B. Meetings, Pausen). Die Therapeuten füllten im Wesentlichen ein Instrument – das Resident Assessment Instrument – Mental Health (RAI-

MH) aus, kodierten den Medikamentenkonsum, diagnostische Leistungen und Elektrokrampftherapiebehandlungen über einen Zeitraum von 7 Tagen. Allerdings gingen die diagnostisch erbrachten Leistungen und die Elektrokrampftherapiebehandlungen nicht in die Berechnung der Kostenvariablen ein. Des Weiteren wurden die Kosten der Medikamente und Fixkosten (wie z.B. Verwaltungsvorgänge oder Mahlzeiten) aus der Kalkulation des Ressourcenverbrauches ausgeschlossen. Die Daten wurden mittels eines Entscheidungsbaumverfahrens ausgewertet, wobei nur die Prädiktoren, die sowohl nach den Kriterien der statistischen Varianzreduktion als auch nach klinischer Plausibilität, eingeschätzt durch Experten (Psychologen, Krankenschwestern und Psychiater), für die Bildung der Gruppen ausgewählt wurden. Nach Expertenworkshops und einer abschließenden Konsensbildung wurde das endgültige Modell für den Case-Mix Algorithmus festgelegt.

Das Klassifikationsmodell beinhaltete einen 7-stufigen hierarchischen Gruppierungsalgorithmus, welcher die Patienten ausgehend von ihrer Erstdiagnose zu insgesamt 47 Fallgruppen zuordnet. Die 47 Fallgruppen erklären in der Studie von Hirdes et al. (2003) 26,3% der Gesamtvarianz der durchschnittlich täglich anfallenden Kosten. Abbildung 6 stellt einen Ausschnitt für Patienten mit affektiven Störungen des SCIIP Fallgruppensystems dar. Demnach ist eine weitere Aufsplittung der Patienten dieser Diagnosegruppe zunächst von der „Aufenthaltsdauer in Tagen“ bestimmt. Liegt die Verweildauer der Patienten unter 4 Tagen spielen noch die Variablen „Vorhandensein psychotisch/affektiver Symptome“ und „Suizidalität“ bei der Zuordnung zu einer Fallgruppe eine Rolle. Beträgt die Verweildauer der Patienten allerdings über 5 Tage tragen Variablen wie z.B. Suizidrisiko, kognitive Leistungsfähigkeit oder Gesundheitsprobleme zur Trennung der Gruppen bei. Weitere klinische Variablen, die signifikant Variation im Ressourcenverbrauch aufklärten, waren z.B. Verhaltensauffälligkeiten, Fremdgefährdung, forensicher Status oder Einsichtsfähigkeit.

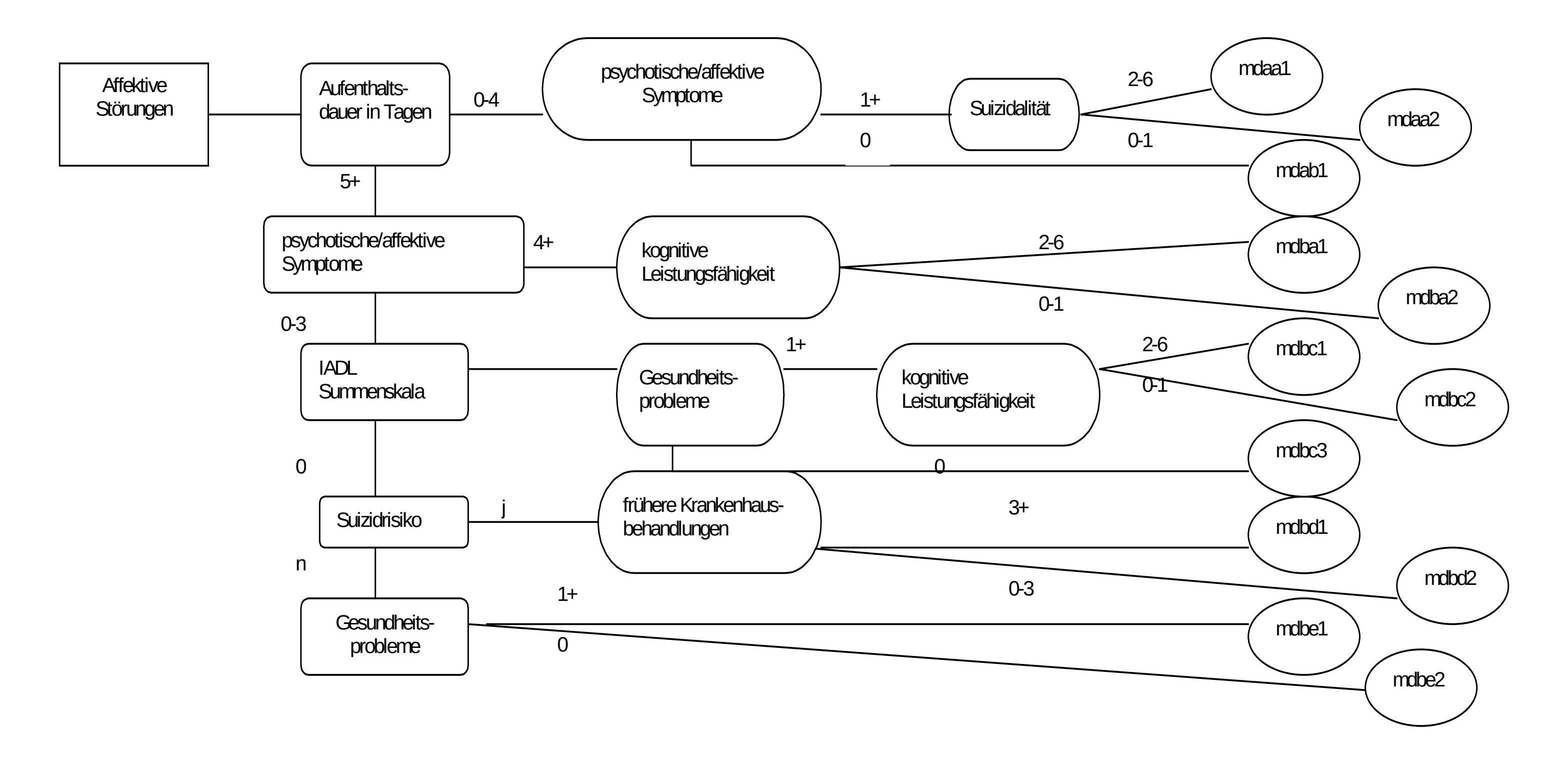



Abbildung 6: Entscheidungsalgorithmus für die Diagnosegruppe „Affektive Störungen" des SCIIP-Systems von Hirdes et al. (2003).
*mdaa1 bis mdbe2 sind die Bezeichnungen der Fallgruppen; n = nein; j = ja; IADL = Instrumental Activities of Daily Living

Bei einer begrenzten Anzahl von Patienten wurde auch die Aufenthaltsdauer zur Bestimmung der Gruppen herangezogen, weil die vorausgegangene bivariate Analyse überraschend einen substantiellen Unterschied zwischen Aufenthaltsdauer und täglichem Ressourcenverbrauch aufzeigte. Patienten der ressourcenintensivsten Gruppe waren Patienten mit einer Schizophrenen Störung, die eine Aufenthaltsdauer zwischen 0 bis 4 Tagen und eine oder mehrere Verhaltensauffälligkeiten aufwiesen. Dagegen gehörten Patienten mit einer Schizophrenen Störung und langer Behandlungsdauer (mehr als 750 Tage), keiner Beeinträchtigung im Funktionsniveau (gemessen mit dem Activity of Daily Living, ADL), aber moderaten bis höheren Scores bei psychotischen und affektiven Symptomen zur ressourcengünstigsten Gruppe. Überraschend bei diesem Befund ist, dass Patienten mit kürzeren Aufenthaltsdauern mehr Ressourcen verbrauchen als Patienten die langfristig behandelt werden. Möglicherweise könnte das in einer intensiveren medizinischen oder pflegerischen Betreuung (z.B. 24 Stunden Fixierung mit ständiger Betreuung) begründet sein.

Das entwickelte System bietet nach Meinung der Autoren (Hirdes et al., 2003) einige Vorteile. Zum einen erfüllt es die Ansprüche an Praktikabilität, da es sich auf Informationen, die zu

Beginn der Behandlung mit dem RAI-MH ausgefüllt werden können, gründet. Zum anderen ist es relativ robust gegen den ökonomischen Anreiz, kürzere Verweildauern zu schaffen, da es sich auf Tageskosten bezieht. Eine weitere Besonderheit dieses Systems ist der weitest gehende Verzicht auf Servicevariablen, Klinikmerkmale und „verfälschbare“ Kriterien (wie z.B. der GAF-Score, Hirdes et al., 2003). Hirdes et al. (2003) führen aber auch an, dass in einigen Fällen die Gruppen aufgrund einer zu geringen Fallzahl nicht weiter aufgeteilt werden konnten, obwohl der Varianzkoeffizient höhere Werte als erwartet aufwies. Genauso verhielt es sich auch mit einigen Störungsgruppen (z.B. Essstörungen oder Persönlichkeitsstörungen), so dass die Autoren zu dem Schluss kommen, dass weitere Forschung mit vorab höheren Fallzahlen für diese Gruppen nötig sein wird.

### *3.2.6 Die Patient Management Categories (PMC)*

Dieses Patientenklassifikationssystem wurde vom Pittsburgh Research Institute in Kooperation mit **amerikanischen Ärzten für den stationären Krankenhausbereich** entwickelt. Die Patient Management Categories (PMCs) bestehen aus 47 Modulen, die 830 nach Aufnahmegrund, Diagnose und Haupttherapie gegliederte Patientengruppen beinhalten (Young et al., 1985). Psychiatrische Diagnosegruppen befinden sich in den Modulen von 4501 bis 4518. Einem Behandlungsfall dürfen mehrere Module gleichzeitig zugeordnet werden, wobei aber aus jedem Modul nur eine Patientengruppe ausgewählt wird (Fischer, 1997). Innerhalb eines Moduls sind deshalb die Patientengruppen nochmals in Schweregrade unterteilt. Die PMCs wurden aufgrund von Expertenpanels gewonnen, die zusätzlich die Aufgabe hatten, die zu erwartenden therapeutischen Leistungen eines prototypischen Patienten in einer PMC zu definieren. Daraus folgend wurden Behandlungspfade (Patient Management Paths) gebildet, die sich über *Aufnahmegrund*, *diagnostische Leistungen*, *Diagnosen* und *therapeutische Leistungen* definieren und zur Berechnung der Behandlungskosten führen.

Im Vergleich zur Klassifikation von Patienten auf der Basis der DRGs wird mit den PMCs ein leicht höherer Varianzanteil an der Aufenthaltsdauer erzielt (15% vs. 18%) (Fischer, 1997).

### 3.2.7 Alternate DRGs (ADRGs)

Taube et al. (1984b) konstruierten unter Nutzung weiterer Kriterien alternative DRG-Gruppen an einer Gesamtstichprobe von N = 12.618 in Ergänzung zu den Fallgruppen des damals vorgehaltenen amerikanischen DRG-Systems für Patienten mit psychischen Störungen.

Es wurden 35 DSM-III Diagnosegruppen hinsichtlich der Homogenität der Behandlungsdauer in 6 Kategorien unterteilt. Abbildung 7 zeigt exemplarisch die Gruppierung am Beispiel der Diagnose „Major Depression". Dabei spielt zur weiteren Aufteilung dieser Gruppe der Einweisungsstatus, Behandlungsmaßnahmen (z.B. Entgiftung ohne Training) und das Alter der Patienten eine bedeutsame Rolle. Unter Subsummierung folgender Kriterien: *Art der Behandlungsmaßnahme*, *Alter*, *Familienstand*, *Einweisungsmodus*, *Entlassungsart*, *vorherige therapeutische Versorgung* und *rechtlicher Status* entstanden schließlich 22 alternative DRG-Gruppen (ADRG).

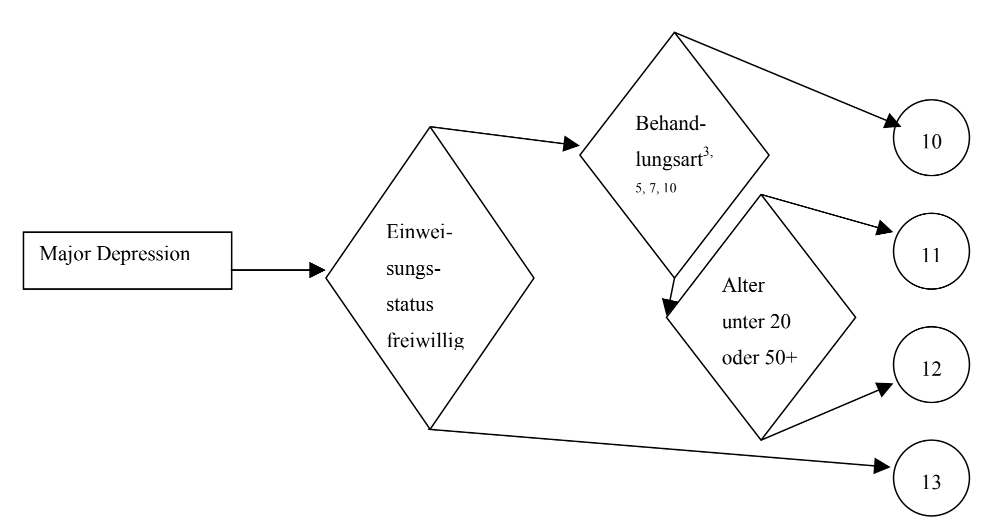



Abbildung 7: Gruppierungsprozess des ADRG-Systems von Taube et al. (1984) für die Diagnose Major Depression (DSM-III basierend) in Anlehnung an Burgmer et al. (2003).

*10 bis 13 ist die Fallgruppenzuordnung, Behandlungsarten: Entgiftung ohne Training, keine verbale Therapie ohne Training, individuelle Therapie ohne Training, individuelle und Gruppentherapie ohne Training

Die so gebildeten Gruppen des ADRG-Systems klärten insgesamt 11,8% der Varianz der Behandlungsdauer im Gegensatz zu den Fallgruppen des amerikanischen DRG-Systems mit 3,2% Varianzaufklärung auf. Unter Einbezug struktureller Merkmale (Art des Krankenhauses und Region) konnten die ADRG-Gruppen 21,4% der Varianz der Behandlungsdauer erklären.

#### *3.2.8 Systemetrics´s Disease Staging (SDS) und Clinically Related Groups (CRG)*

Mitchell et al. (1987) verglichen drei verschiedene Klassifikationssysteme hinsichtlich ihrer Eignung zur Abbildung des Ressourcenverbrauches. Grundlage bildete eine Stichprobengröße von mehr als 22.000 Medicare-Versicherten mit psychischen Störungen, die sich zur stationären Behandlung in Allgemeinkrankenhäusern aufhielten (Mitchell et al., 1987). Der Vergleich bezog sich auf die Fallgruppen des amerikanischen DRG-Systems, auf **Systemetrics´s Disease Staging (SDS)** und auf von den Testautoren entwickelten **Clinically Related Groups (CRG)**.

Das Prinzip des **SDS Systems** beruht auf einer Unterteilung aller Erkrankungen in vier verschiedene Krankheitsstufen: 1) Erkrankungen ohne Komplikationen oder mit Problemen niedriger Schwere, z.B. eine schizophrene Störung 2) bezogen auf ein Organ mit beträchtlich erhöhtem Komplikationsrisiko, z.B. hirnorganisches Syndrom mit Psychose 3) mehrere Körperbereiche betreffend mit genereller Organbeteiligung und schlechter Prognose, z.B. senile Demenz vom Typus Alzheimer und 4) Tod. Der Algorithmus läuft über alle Diagnosen und ordnet diese einem entsprechenden Krankheitsstadium zu. Somit entstanden ca. 1000 Kategorien.

Mit dem **CRG System** versuchten die Testautoren die amerikanischen Fallgruppen für den Bereich psychischer Störungen zu überarbeiten und klinisch bedeutungsvolle Gruppen zu entwickeln, die aber auch an der Diagnose orientiert waren. Zur Unterteilung benutzten Mitchell et al. (1987) zusätzlich die Variable Alter. Allerdings wurden auch einige der ursprünglichen Fallgruppen des amerikanischen DRG-Systems unmodifiziert übernommen. Letztendlich entstanden so 16 klinisch bedeutungsvolle Gruppen (CRGs). Ein Beispiel für eine CRG ist für die Diagnose Persönlichkeitsstörungen zu nennen, der noch die Diagnose Essstörungen zugeordnet wurde.

Bezüglich der Kosten der Krankenhausbehandlung klärten die Fallgruppen des amerikanischen DRG-Systems im Durchschnitt über die 4 beteiligten Bundesstaaten 4,1% der Varianz auf, die SDS-Gruppen 7% und die CRGs 6%. Bei der Varianzreduktion der Verweildauer konnten leicht höhere Werte erzielt werden: amerikanisches DRG-System 7%, SDS 9,6% und CRGs 9,3%. Die Varianzkoeffizienten variierten dabei sehr stark zwischen den einzelnen

Bundesstaaten (z.B. für CRGs in Michigan 11,2%, New Jersey 5,4%, North Carolina 8,4% und Washington 12,3%).

### 3.3 Dimensionen des Ressourcenverbrauchs und Ableitung potentieller Prädiktoren zur Fallgruppenbildung für Patienten mit psychischen Störungen

Ein Klassifikationssystem für den Bereich psychischer Störungen sollte Patienten hinsichtlich **klinisch relevanter** Merkmale zu Gruppen zusammenfassen und Fallgruppen bilden, die sich in ihrem Ressourcenverbrauch gleichen. Des Weiteren sollte es den Kriterien der **Transparenz**, der **Praktikabilität**, der **Reliabilität** und der **Validität** entsprechen (Fischer, 1997). Damit stellt sich die Aufgabe, klinisch relevante Dimensionen abzuleiten, ein Instrument zur Erfassung von Fallgruppen mit möglichst homogenem Ressourcenverbrauch zu konzipieren und das entwickelte System hinsichtlich statistischer und psychometrischer Gütekriterien zu testen. Im Rahmen dieser Aufgabe und unter dem Gesichtspunkt der Praktikabilität ist es ebenso essentiell, diejenige Dimension des Ressourcenverbrauchs und die damit verbundenen Prädiktoren zu identifizieren, die eine möglichst hohe Varianzaufklärung leistet.

Zur Ableitung von Klassifikationsdimensionen für Fallgruppen im psychischen Bereich erscheint das **MH-CASC Projekt** (Buckingham et al., 1998), das **NZ-CAOS Projekt** (Gaines et al., 2003), der **PSI Index** (Horn et al., 1989), das **SCIIP System** (Hirdes et al., 2003) und in einigen Bereichen auch das **PPC-System** geeignet (Ashcraft et al., 1989). In allen Projekten lag die aufgeklärte Varianz für die Verweildauer und den Ressourcenverbrauch bei ca. 16 bis 50%, was zunächst für eine klinische Relevanz der zugrunde liegenden Kriterien spricht, was aber auch Hinweise darauf gibt, dass noch weitere Dimensionen den Ressourcenverbrauch substantiell mitbestimmen. Dabei erreichten zwei Projekte (PSI Index, Horn et al., 1989 und SCIIP System, Hirdes et al., 2003) die höchsten Varianzaufklärungskennwerte (vgl. hierzu auch Tabelle 2). Da es sich bei diesen beiden Studien allerdings um die Konstruktion von retrospektiven Patientenklassifikationssystemen handelt und die Fallgruppen auf der Basis von bereits für den Ressourcenverbrauch relevanten Merkmale (z.B. Pflegeaufwand oder Aufenthaltsdauer in Tagen) entwickelt wurden, sollen die weiteren Kriterien dieser beiden Projekte nur bedingt zur Ableitung von Klassifikationsdimensionen eines nationalen Systems herangezogen werden. Daraus kann aber auch geschlussfolgert werden, dass vermutlich die hohen Varianzaufklärungskennwerte des Ressourcenverbrauches dieser beiden Projekte durch die Aufnahme der unmittelbar den Kosten zugrundeliegenden Variablen zu erklären sind.

Die PMCs schlagen durch die Definition von Behandlungspfaden eine Brücke zwischen Klassifikationskriterien und therapeutischen Leistungen (Young et al., 1985). Dabei weisen sie allerdings eine zu hohe Anzahl an Patientenkategorien auf (870 Kategorien), so dass das System aufgrund mangelnder Praktikabilität für eine weitere Analyse der zugrunde liegenden Kriterien nicht in Betracht gezogen werden soll. Entsprechend verhält es sich mit den von Taube et al. (1984a, 1984b) und Mitchell et al. (1987) entwickelten Systemen (ADRGs, CRG und SDS). Die Varianzaufklärungskennwerte dieser Systeme lagen allerdings in einem Bereich, der auch mit den bereits etablierten DRG-Systemen zu erzielen ist, weswegen sie nicht weiter berücksichtigt werden sollen. Ganz im Gegensatz zu den gerade aufgeführten Systemen verfolgte das MHG Projekt von Carthew et al. (2003) den Ansatz, Fallgruppen primär auf der Grundlage von Interventionen zu bilden (retrospektives Design). Da dieses Vorgehen von dem methodischen Zugang vorliegender Arbeit, welches prospektiv angelegt ist, sehr abweicht, soll es von einer näheren Überprüfung ausgeschlossen werden.

Erste Entwicklungen zu zugrunde liegenden Dimensionen des Ressourcenverbrauches in der medizinischen Rehabilitation für Patienten mit psychischen Störungen finden sich bei Koch et al. (1998). Innerhalb des Projektes wurde auf der Basis von Therapieplänen eine Ableitung von Klassifikationsdimensionen für Hauptbehandlungsgruppen u.a. auch im psychischen Bereich vorgenommen. Unter Berücksichtigung der bereits vorliegenden Ableitungen von Koch et al. (1998) und den Dimensionen der international entwickelten Klassifikationssysteme MH-CASC Projekt, NZ-CAOS, PSI, SCIIP und PPC könnten sich folgende Faktoren als zentrale Klassifikationsdimensionen anbieten:

1. Krankheitsbezogene Faktoren
2. Patientenbezogene Faktoren
3. Dimension des Therapieverlaufs.

Bezüglich der **krankheitsbezogenen Faktoren** haben sich die folgenden vier Gruppen von Variablen empirisch als besonders bedeutsam erwiesen: (1) *Psychische und somatische Haupt- und Nebendiagnose*, (2) *Symptombelastung, Schweregrad und Erkrankungsstadium*, (3) *bisheriger Krankheitsverlauf* und (4) *Status der Einweisung*. Dabei kommen den vier Gruppen jeweils unterschiedliche Bedeutungen zu. Trotz der Varianz innerhalb der Diagnosegruppen fällt auf, dass die Diagnose in allen Klassifikationssystemen als gesetztes Merkmal

gilt und deswegen immer wieder zur Fallgruppierung herangezogen wird. Auch die Variablen Symptombelastung, Schweregrad, Erkrankungsstadium und bisheriger Krankheitsverlauf erweisen sich in den aufgeführten Studien als besonders gruppierungsrelevant, wobei allerdings kein Konsens hinsichtlich ihrer genauen Operationalisierung besteht (vgl. dazu Kapitel 5). Obwohl die Vorstellungen zur Verwendung von Instrumenten zur Erfassung des Schweregrades auseinander gehen, erwies sich ein Instrument, die „Health of the Nation Outcome Scales, HoNOS“, innerhalb von drei Projekten (MH-CASC, NZ-CAOS und MHG) hinsichtlich der Fallgruppenbildung als ein bedeutsamer Prädiktor des Ressourcenverbrauchs. Eine ebenso wichtige Variable für den Ressourcenverbrauch scheint der Status der Einweisung (z.B. Notaufnahme, freiwillige Einweisung) der Patienten in die stationäre psychiatrische/psychotherapeutische Behandlung zu sein. Da in der stationären Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen die Zuweisung gesteuert ist und z.B. keine Notaufnahmen vorkommen, wird diese Variable für ein Klassifikationssystem in diesem Bereich keine Rolle spielen und soll hier nicht weiter berücksichtigt werden.

Hinsichtlich **patientenbezogener Faktoren** haben sich in den vorliegenden Studien insbesondere die beiden Variablen *Alter* und *soziale Unterstützung* als Kriterien zur Fallgruppenbildung als geeignet erwiesen. So ist z.B. das Merkmal Alter der Patienten in dem Gruppierungsmodell von Buckingham et al. (1998) der bedeutsamste Prädiktor zur Trennung der Gruppen. Bei der Variable soziale Unterstützung dagegen könnte es sich um eine Proxyvariable des Familienstandes der Patienten handeln. Entgegen theoretischer Überlegungen zeigten sich jedoch Variablen wie z.B. Schulbildung, Geschlecht oder Berufsstand für eine Fallgruppierung als weniger bedeutsam. So könnte doch z.B. davon ausgegangen werden, dass ein arbeitsloser Patient als Therapieziel die Wiedereingliederung ins Arbeitsleben benennt und im Therapieverlauf mehr Ressourcen in Form von sozialarbeiterischen Tätigkeiten anfallen würden.

Eine dritte Dimension, die grundlegend sein könnte, Behandlungsfälle zu Fallgruppen mit möglichst homogenem Ressourcenverbrauch zusammenzufassen, bezieht sich auf den **Therapieverlauf**. Diejenigen Variablen aus den vorliegenden Patientenklassifikationssystemen, die sich zur Gruppierung als bedeutsam herausgestellt haben, waren *Komplikationen während des Aufenthaltes*, *Therapiecompliance*, die *Einschätzung der Behebung von akuter Symptomatik* und die *Entlassungsfähigkeit*. Unter den eben aufgeführten Merkmalen ist dabei das Kriterium „Komplikationen während des Aufenthaltes“ hervorzuheben, wobei auch hier unter-

schiedliche Ansätze zur Operationalisierung bestehen. So fassen Horn et al. (1989) unter Komplikationen lediglich unerwünschte Ereignisse, die negative Auswirkungen auf den somatischen Zustand des Patienten haben (z.B. Drogenintoxikation). Hirdes et al. (2003) subsumieren darunter auch Variablen, wie z.B. Vorhandensein von Suizidalität und Suizidrisiko. Da es u.a. Zielsetzung vorliegender Arbeit ist, ein Patientenklassifikationssystem zur Vorhersage des Ressourcenverbrauches zu entwickeln, sollen diese Kriterien, die retrospektiv gewonnen wurden, nicht weiter berücksichtigt werden.

Es bleibt anzumerken, dass es sich bei o.g. Kriterien um Variablen handelt, die international und damit über verschiedene Kulturkreise bzw. unterschiedliche Versorgungsstrukturen hinweg als gruppierungsrelevant ermittelt wurden. So zeigte sich z.B. die Relevanz der HoNOS in verschiedenen Versorgungssystemen psychisch Kranker in Australien, Neuseeland und Großbritannien. Daher bleibt zu überprüfen, ob sich die in diesem Kapitel aufgeführten Kriterien im deutschen Kulturkreis und in dem Versorgungsbereich der stationären Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen wieder finden und bewähren.

Seit einigen Jahren verstärken sich Bemühungen von nationalen Forschungs- und Arbeitsgruppen, die Determinanten des Ressourcenverbrauches von psychisch Erkrankten zu untersuchen. Da es bislang noch kein differenziert ausgearbeitetes nationales Klassifikationssystem für Patienten mit psychischen Störungen gibt, wurde eine Überprüfung der Arbeiten, die sich mit Einflussfaktoren auf Ressourcenverbrauch/Behandlungsdauer befassen, vorgenommen und diese auf Relevanz für die Definition nationaler Kriterien des Ressourcenverbrauches untersucht. Das folgende Kapitel soll einen Überblick über die empirischen Befunde nationaler Studien geben und nationale Ansätze zur Erfassung des Ressourcenverbrauches darstellen.

# 4 Nationale Ansätze zu Prädiktoren des Ressourcenverbrauchs von Patienten mit psychischen Störungen

Im Folgenden soll geprüft werden, in wie weit sich die empirischen Befunde nationaler Studien zum Ressourcenverbrauch den identifizierten Dimensionen (vgl. Kap.3, Abschnitt 3.3) zuordnen lassen und in wie weit Übereinstimmungen bzw. Unterschiede mit den Klassifikationskriterien internationaler Systeme bestehen. Der Systematisierung folgend sollen die empirischen Befunde zur krankheitsbezogenen und zur patientenbezogenen Dimension sowie zur Dimension des Therapieverlaufs dargestellt werden.

## 4.1 Nationale, empirische Befunde zur krankheitsbezogenen Dimension

Verschiedene Studien aus unterschiedlichen Versorgungsbereichen von Patienten mit psychischen Störungen untersuchten die Fragestellung, in wie weit **krankheitsbezogene Merkmale** in Zusammenhang mit dem Ressourcenverbrauch in stationärer Therapie stehen (Scheytt et al., 1996; Zielke et al., 1997; Jerschke et al., 1998; Borgart & Meermann, 1999; Kächele für die Studiengruppe MZ-ESS, 1999; Richter, 1999, 2001; Kluge et al., 2002; Krautgartner et al., 2002; Mestel et al., 2002; Spießl et al., 2002; Spitzer et al., 2002; Roick et al., 2002a; Roick et al., 2002b). Wie bereits im Kapitel 3, Abschnitt 3.3 zu Kriterien internationaler Klassifikationssysteme dargelegt, zeigt sich zunächst eine große Variation der abgeleiteten Merkmale und deren Operationalisierungen. So werden z.B. Variablen wie die Diagnose von Patienten, der Schweregrad, die Dauer der Erkrankung, die Arbeitsunfähigkeitsdauer und der Einweisungsmodus in die krankheitsbezogene Dimension miteinbezogen.

Allgemein lässt sich feststellen, dass derzeit zahlreiche Befunde aus dem Versorgungsbereich der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie existieren, dagegen aber bislang kaum empirische Studien aus der Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie und der Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen vorgelegt worden sind. Tabelle 5 vermittelt einen Überblick über zentrale nationale Studien, die den Einfluss verschiedener Faktoren auf die Behandlungsdauer in Deutschland untersucht haben. Im Folgenden sollen die zentralen Ergebnisse der in der Tabelle 5 zusammengetragenen Studien näher aufgeführt werden.

Tabelle 5: Überblick über zentrale nationale Studien zu Einflussfaktoren auf die Behandlungsdauer im Bereich psychischer Störungen

| *Studie* | *Anwendungsbereich* | *Stichprobengröße* | *Erhobene Variablen und eingesetzte Instrumente* | *Operationalisierung des Ressourcenverbrauches* | *Anzahl der Fallgrupppen /Varianzaufklärung* |
|---|---|---|---|---|---|
| **Studien zu Einflussfaktoren auf die Behandlungsdauer (Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie und Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen)** | | | | | |
| Borgart & Meermann (1999) | Patienten mit Angststörungen | 1.173 | Basisdokumentation | Behandlungsdauer | nicht berechnet |
| Kächele (1999) | Patienten mit Essstörungen | 1.171 | keine Angabe | Behandlungsdauer, Kosten | 2-8% |
| Bassler et al. (1995) | Patienten einer Klinik für Psychosomatik und Psychotherapie | 219 | Klinische und therapeutische Merkmale, GBB, Gießen-Test, HAQ, STAI, Einschätzung des Therapieerfolgs | Behandlungsdauer | keine Angabe |
| Zielke et al. (1997) | Patienten psychosomatischer verhaltenstherapeutischer Fachkliniken | 22.838 | Basisdokumentation | Behandlungsdauer | keine Angabe |
| Mestel et al. (2002) | Psychosomatische Fachklinik | 1.928 | Basisdokumentation; OPD, SCL-90R, IIP, IRES | Behandlungsdauer | Keine Angabe |
| Von Heymann et al. (2003) | 17 Fachkliniken für die Behandlung von Patienten mit psychischen Störungen | 17.860 | Basisdokumentation | Behandlungsdauer | k.A. |
| **Studien zu Einflussfaktoren auf die Behandlungsdauer (Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie)** | | | | | |
| Roick et al. (2002a) | stationärer psychiatrischer Bereich | 969 | AOK-Krankenkassendaten | Behandlungskosten | 3-9% |
| Krautgartner et al. (2002) | Stationärer psychiatrisch/psychotherapeutische Fachkliniken in Österreich | 17.154 | Basisdokumentation | Kumulierte Krankenhaustage | k.A. |
| Richter et al. (1999) | Stationäre psychotherapeutische /psychiatrische Fachklinik | 2.450 | Basisdokumentation | Behandlungsdauer | 10,5-11,3% |
| Spießl et al. (2002) | stationärer psychiatrischer Bereich | 1.811 | Basisdokumentation | kumulierte stationäre Behandlungsdauer | 10,3 - 17,4% |

### *4.1.1 Zusammenhang zwischen Diagnose und Ressourcenverbrauch*

In der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie rückt seit einiger Zeit vor allem eine Patientengruppe in den Forschungsfokus. Diese Patientengruppe, auch **heavy user** genannt, ist durch einen hohen Ressourcenverbrauch bzw. eine hohe Inanspruchnahme therapeutischer Leistungen gekennzeichnet. Dabei wird geschätzt, dass der Anteil der heavy user bei ca.

10% bis 30% aller Patienten in der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie liegt (Roick et al., 2002b). Die Festlegungen, welche Kriterien erfüllt sein müssen, damit ein Patient zur Kategorie des „heavy users“ zählt, sind sehr unterschiedlich. Häufig sind in der Definition die Zahl der stationären Aufnahmen (meist 1 bis 3 Krankenhausaufenthalte pro Jahr) eingeschlossen (Roick et al., 2002b). Einige Autoren berücksichtigen allerdings auch die Summe der Behandlungstage im Beobachtungszeitraum oder die verursachten Versorgungskosten. Im Weiteren soll geprüft werden, ob zwischen den Merkmalen dieser Patientengruppe Gemeinsamkeiten mit der Gruppe der verstärkt Ressourcen in Anspruch nehmenden Patienten in stationärer psychosomatischer/psychotherapeutischer Behandlung bestehen und welche Diagnosen sich als ressourcenverbrauchsrelevant herauskristallisiert haben.

Betrachtet man die Ergebnisse der heavy-user-Studien **aus der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie,** die zum Zusammenhang zwischen der **Diagnose** eines Patienten und dem Ressourcenverbrauch bzw. der Behandlungsdauer durchgeführt wurden, so lässt sich weitest gehend konsistent feststellen, dass vor allem das Vorhandensein der Hauptdiagnose „Schizophrene Störung“ und/oder „Persönlichkeitsstörung“ mit einer höheren Inanspruchnahme an Krankenhausleistungen verbunden war (Krautgartner et al., 2002; Spießl et al., 2002; Roick et al., 2002a; Roick et al., 2002b). Dagegen traten die Hauptdiagnosen „Angst-, Belastungs- und somatoforme Störungen“ und „Alkohol- und Drogenabusus“ signifikant häufiger in der Gruppe der Patienten auf, die z.B. durch weniger als drei Krankenhausaufenthalte im Jahr gekennzeichnet sind („ordinary user“). Auch in der Studie von Richter (Richter, 1999) konnte gezeigt werden, dass die Hauptdiagnose einer Suchterkrankung oder einer „Angst-, Belastungs- und somatoformen Störung“ mit kürzeren Verweildauern einher ging.

Auch in der **Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie** und in der **Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Stör**ungen sind die empirischen Befunde zum Zusammenhang zwischen **Diagnose** und Ressourcenverbrauch bzw. Behandlungsdauer weitestgehend konsistent. Es zeigte sich, dass Patienten mit der Hauptdiagnose „Borderline Persönlichkeitsstörung“, „Zwangsstörung“ und „Essstörung“ die längsten Behandlungsdauern aufwiesen (Zielke et al., 1997; Mestel et al., 2002; von Heymann et al., 2003). Die Patienten der Diagnosegruppe „Essstörungen“ kann dahingehend differenziert werden, das Patienten mit einer „Anorexia nervosa“ einen signifikant höheren Ressourcenverbrauch (Behandlungskosten) aufwiesen als Patienten mit einer „Bulimia nervosa“ (Kächele für die Studiengruppe

MZ-ESS, 1999). Signifikant kürzere Verweildauern wurden für Patienten mit einer „Depressiven Störung“ und „Angst- und Anpassungsstörung“ berichtet (Mestel et al., 2002).

Als **Fazit** lässt sich festhalten, dass die empirischen Befunde aus der heavy-user-Forschung auf den Bereich der psychosomatischen/psychotherapeutischen Behandlung nur schwer übertragbar sind. Die Ergebnisse zeigen konsistent, dass das Vorliegen einer „Schizophrenen Störung“ mit einer höheren Inanspruchnahme an Krankenhaustagen und -aufenthalten verbunden ist. Da allerdings in dem Versorgungsbereich der Rehabilitation von Patienten mit psychischen/psychosomatischen Störungen Diagnosen aus dem schizophrenen Formenkreis nur mit ca. 5% vertreten sind (Schulz et al., 2004), sind die Ergebnisse nicht als unmittelbar zentral für die vorliegende Untersuchung zu werten. Die Befunde aus dem Versorgungsbereich der Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen und der Krankenhausbehandlung sind allerdings bedeutsam für die vorliegende Arbeit. Hier sollte überprüft werden, ob sich die Diagnoseoberkategorien (z.B. Essstörungen) durch eine differenzierte Schweregraderfassung noch weiter hinsichtlich des Ressourcenverbrauchs aufschlüsseln lassen.

Betrachtet man das Kriterium „Hauptdiagnose“ zur Fallgruppenbildung kritisch, so besteht **ein weiteres Problem** darin, dass eine Vielzahl an Studien eine mangelnde Reliabilität und Validität der Diagnosestellung nachweisen konnten. Eine aktuelle Untersuchung von Keller (2000) zur Reliabilität und Validität des F-DIPS (Diagnositsches Interview bei psychische Störungen, Forschungsversion für DSM-IV von Margraf et al., 1996) anhand einer Stichprobe von N = 191 Patienten zeigte, dass sehr gute Retest-Reliabilitäten (Intervall: 2 Wochen) für die Diagnoseerstellung einer Essstörung, Substanzmissbrauch, Major Depression, bipolaren Störung und Hypochrondie bestanden. Mangelnde Retest-Reliabilitäten fanden sich vor allem im Bereich der Angststörungen (z.B. Panikstörung ohne Agoraphobie) und der somatoformen Schmerzstörung. Hinsichtlich der Validität zeigten sich zudem Defizite in den Bereichen der Diagnostik einer Angststörung, Affektiven Störung und Somatoformen Störung. Zieht man in Betracht, dass diese Ergebnisse unter standardisierten Interviewbedingungen gewonnen wurden, was in der klinischen Praxis aber oft aus ökonomischen Gründen nicht realisierbar erscheint, so stimmen diese Ergebnisse hinsichtlich der Hauptdiagnose als reliables und valides Kriterium bedenklich. Nichtsdestotrotz gilt die Diagnose bei der Fallgruppenbildung, wie in Kap. 3 bereits dargelegt, als gesetztes Merkmal.

#### *4.1.2 Zusammenhang zwischen Schwere der Erkrankung und Ressourcenverbrauch*

Während hinsichtlich des Merkmals „Diagnose“ unter methodischen Gesichtspunkten oft Probleme mit der Reliabilität und Validität bestehen, stellt sich bei der Erfassung des Schweregrades einer Erkrankung eher das Problem der Heterogenität der Operationalisierung. Kritisch ist deswegen anzumerken, dass bislang kein einheitliches Modell zum Schweregrad einer psychischen Erkrankung existiert und damit völlig unklar bleibt, welche Merkmale essentiell für eine Schweregradbestimmung wären.

Generell lassen sich die Befunde wieder grob in empirische Ergebnisse aus der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie und Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie sowie Rehabiliationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen einteilen. Da der folgende Abschnitt einen zentralen Aspekt der vorliegenden Arbeit berührt, sollen die entsprechenden Studien ausführlicher vorgestellt werden.

1) *Empirische Befunde zur Schwere einer Erkrankung aus der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie*

Auf der Basis eines **Literaturreviews** identifizierten Roick et al. (2002b) Merkmale des Schweregrades bei Patienten mit einer hohen Inanspruchnahme stationärer Versorgungsleistungen (sog. heavy user, siehe dazu auch Abschnitt 4.1.1). Sie konnten aufzeigen, dass diese signifikant mehr stationäre Voraufnahmen als Normalnutzer aufwiesen (3 Studien), signifikant weniger Krankheitseinsicht besaßen (2 Studien), deutlich weniger compliant bei der Medikamenteneinnahme (5 Studien) oder bei anderen Therapiemaßnahmen waren (3 Studien) und signifikant mehr selbst- oder fremdgefährdet waren. Das Merkmal „stationäre Voraufnahmen“ sagte mit der höchsten Varianzaufklärung eine erneute Krankenhausaufnahme voraus (7 Studien). In einer weiteren **Literaturübersicht** zu Prädiktoren der Behandlungsdauer identifizierte Richter (2001) u.a. folgende Kriterien des Schweregrads: Anzahl früherer Behandlungen bei schizophrenen Patienten, besondere Belastung von Angehörigen depressiver gerontopsychiatrisch behandelter Patienten, Gewalthandlungen der Erkrankten, Krankenhaustyp, Kostenträger bzw. Klinikcharakteristika, Komorbidität, PPSI Schweregradindex von Horn et al. (1989) Schwere der Symptomatik, Arbeitsintegration und soziale Unterstützung sowie soziodemographische Merkmale (siehe dazu Kap. 3, Abschnitt 3.3). Diese Kriterien klärten über 20% der Varianz der Behandlungsdauer auf (7 Studien).

Richter (1999) führte eine **empirische Untersuchung** zum Zusammenhang des Schweregrads eines Patienten und der Behandlungsart auf der Grundlage der Psychiatrie-Personalverordnung (Psych-PV, Kunze & Kaltenbach, 1996) an 178 Patienten einer allgemeinpsychiatrischen Krankenhausabteilung durch. Zunächst teilten die Autoren die Patienten unterschiedlichen Kategorien der Psych-PV[6] zu. Anschließend verglichen sie diejenigen Patienten, die in die Kategorie A1 „Regelbehandlung" eingestuft worden waren, mit Patienten der Kategorie A2 „Intensivbehandlung". Der Schweregrad wurde in dieser Studie mit der Brief Psychiatric Rating Scale (BPRS, Overall & Gorham, 1976), der Clinical Global Impressionscale (CGI, National Institute of Mental Health, 1976) und der Global Assessment of Functioning Scale (GAF, American Psychiatric Association, 1996) erhoben (siehe dazu Kap. 5, Abschnitt 5.1.3). Des Weiteren wurde zur Erfassung der gesundheitsbezogenen Lebensqualität die Short-Form 36 (SF-36, Bullinger & Kirchberger, 1998) eingesetzt. Die Patienten der A2 Kategorie wiesen im CGI, in den Subskalen „Hostilität", „Aktivierung" und „Anergie" der BPRS einen signifikant höheren Belastungsgrad als die Patienten der A1 Kategorie auf. In der GAF Skala zeigten sie geringere Werte als die Patienten, die zur A1 Kategorie zugeordnet wurden, was auf eine höhere Beeinträchtigung hinweist. In der psychischen und somatischen Summenskala der SF-36 ergaben sich keine bedeutsamen Unterschiede zwischen den beiden Gruppen. Signifikant unterschiedlich waren die Patienten der beiden Gruppen auch in Bezug auf den Einweisungsmodus: 42,1% der Patienten in der A2 Kategorie waren zwangseingewiesen im Gegensatz zu 11,8% der Patienten der A1 Kategorie.

---

[6] Die Psychiatrie-Personalverordnung (Psych-PV) unterscheidet drei Kategorien: Allgemeine Psychiatrie, Abhängigkeitskranke und Gerontopsychiatrie. Zur Kategorie „Allgemeine Psychiatrie" gehören: A1 = Regelbehandlung, A2 = Intensivbehandlung, A3 = Rehabilitative Behandlung, A4 = Langandauernde Behandlung Schwer- und Mehrfachkranker, A5 = Psychotherapie und A6 = Tagesklinische Behandlung. Dabei sind weiterhin in der Psych-PV die Minutenwerte pro Woche und Patient für jede dieser Kategorien und für einzelne Berufsgruppen aufgeführt (z.B. Behandlungsbereich A1: 207 Minuten pro Woche und Patient für ärztliches Personal im Gegensatz zu 29 Minuten pro Woche und Patient für psychologisches Personal; Kunze & Kaltenbach, 1996).

2) *Empirische Befunde zur Schwere einer Erkrankung aus der Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie und der stationären Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen*

Anhand einer Stichprobe[7] von N = 1.173 Patienten mit Angststörungen in stationärer Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen untersuchten Borgart und Meermann (1999) folgende **Indikatoren des Schweregrads** in Zusammenhang zur Behandlungsdauer:

- Anzahl stationärer und psychotherapeutischer Vorbehandlungen
- Dauer der Erkrankung und
- Dauer der Arbeitsunfähigkeit vor der Aufnahme.

Dabei zeigte sich in fast allen Ergebnissen ein U-förmiger Trend zwischen dem Schweregrad und der Behandlungsdauer: So steigen die Behandlungszeiten zunächst mit zunehmender Anzahl an stationären und psychotherapeutisch/ambulanten Vorbehandlungen sowie steigender Erkrankungsdauer an, nehmen mit weiter steigender Anzahl der Vorbehandlungen oder längerer Chronifizierung jedoch wieder ab. Die als arbeitsfähig aufgenommenen Patienten wiesen die kürzesten Verweildauern auf.

Weitere Ergebnisse zum Zusammenhang zwischen Schweregrad und Behandlungsdauer lieferte die Studie von Mestel et al. (2002) (siehe Tabelle 5). Die Autoren operationalisierten die Schwere einer Erkrankung über:

- den Beeinträchtigungsschwerescore (BSS, Schepank, 1995)
- die Symptom-Check-List 90R (SCL-90R, dt. Franke, 1995)
- die Konflikt- und Strukturachse der Operationalisierten Psychodynamischen Diagnostik (OPD, Arbeitskreis OPD, 2004)

---

[7] Zur Beschreibung des Settings und der Stichprobe sei an dieser Stelle auf Tabelle 5 verwiesen.

- das Inventar zur Erfassung Interpersoneller Probleme (IIP, Horowitz et al., 1994)
- den Fragebogen zu Indikatoren des Reha-Status (IRES, Gerdes & Jäckel, 1992) und die Arbeitsunfähigkeitszeiten.

Sie fanden heraus, dass die Einschätzung im BSS, die Chronifizierung der Erkrankung des Patienten, die Fremdeinschätzung in der OPD-Strukturachse und die Motivation zur Rehabilitationsbehandlung mit der Verweildauer mit mindestens kleiner Effektstärke korrelierten. Die höchste Korrelation zur Verweildauer wies dabei die Gesamteinschätzung der OPD-Strukturachse auf (r = .42), was bedeutet, dass diejenigen Patienten die fremd eingeschätzt das größte Ausmaß an struktureller Störung (z.B. beeinträchtigte Objektwahrnehmung, Selbstwahrnehmung, Abwehrmechanismen) aufwiesen, auch am längsten behandelt wurden.

Zielke et al. (1997) erfassten die Schwere einer Erkrankung in ihrer Studie (siehe Tabelle 5) über folgende Kriterien:

- Krankheitsdauer
- Dauer der Arbeitsunfähigkeit (AU-Zeit)
- Komorbidität
- Anzahl relevanter Problembereiche für die Behandlung
- Schweregradbeurteilung aus Patientensicht und Vorbehandlungen bis zum Zeitpunkt der stationären psychosomatischen/psychotherapeutischen Rehabilitationsbehandlung.

Im Gegensatz zu Borgart und Meermann (1999) fanden Zielke et al. (1997) eine lineare Beziehung zwischen dem Schweregrad einer Erkrankung eines Patienten und seiner Behandlungsdauer: So stieg der stationäre Behandlungsaufwand mit zunehmenden AU-Zeiten, längerer Chronifizierung, steigender Anzahl an stationären Vorbehandlungen und mit steigender Anzahl behandlungsrelevanter Beschwerden. Weiterhin zeigten die Ergebnisse, dass Patienten, die länger in stationärer Behandlung bleiben, ihre Probleme zu Beginn der Behandlung häufiger schwerer einschätzten als Patienten mit kürzeren Verweildauern.

Auf der Basis einer Analyse von Basisdokumentationen untersuchten von Heymann et al. (2003) u.a. Einflussfaktoren auf die Behandlungsdauer (siehe Tabelle 5). Die Indikatoren des Schweregrads waren wie folgt operationalisiert:

- Anzahl und Dauer der Vorbehandlungen
- Komorbidität
- Chronifizierung und Ausmaß der Beeinträchtigung im BSS.

Ebenso wie Zielke et al. (1997) legen die Ergebnisse von von Heymann et al. (2003) einen linearen Trend zwischen Schweregrad und Behandlungsdauer nahe: So stiegen die stationären Behandlungszeiten signifikant mit der Anzahl und Dauer von Vorbehandlungen, mit zunehmender Anzahl an komorbiden Störungen, mit zunehmender Chronifizierung der Erkrankung und mit höherer Beeinträchtigung im BSS.

In einer weiteren Untersuchung zu Prädiktoren der Dauer stationärer Behandlungen von Patienten mit psychischen Störungen untersuchten Bassler et al. (1995) (siehe Tabelle 5) folgende Merkmale:

- Ausmaß der Belastung im Gießener Beschwerdebogen (GBB, Brähler & Scheer, 1983)
- Ausmaß der Belastung im Gießen-Test (GT-S, Beckmann & Richter, 1983)
- Ausmaß von Angst mit dem State-Trait-Anxiety-Inventar (STAI, Laux et al., 1981)
- Beurteilung des klinischen Störungsgrads mittels einer strukturellen Einschätzung
- psychotherapeutische Vorbehandlungen und Suizidversuche.

Dazu unterteilten sie die Patienten in zwei Gruppen: Patienten mit kürzerer Behandlungszeit (von 36 bis 86 Tagen, n = 104) und Patienten mit längerer Behandlungszeit (von 99 bis 291 Tagen, n = 115). Hinsichtlich des Zusammenhanges zwischen Schweregradeinschätzung durch die Therapeuten und Behandlungsdauer zeigten sich überraschend keine signifikanten Korrelationen. Konsistent mit den Ergebnissen von Zielke (1997) und von Heymann et al. (2003) wiesen auch in dieser Studie die Patienten ohne psychiatrische Vorbehandlungen signifikant kürzere Behandlungsdauern auf. Hinsichtlich psychotherapeutischer Vorbehandlungen und Suizidversuchen in der Kranken-

vorgeschichte zeigten sich nur tendenzielle Ergebnisse, womit ein weiterer Forschungsbedarf verbunden ist: Patienten mit kürzerer Behandlungszeit hatten eher häufiger länger zurückliegende Suizidversuche unternommen, Patienten mit längerer Behandlungsdauer eher aktuelle Suizidversuche. In einer anschließenden Diskriminanzanalyse wurden alle Prädiktoren der Behandlungsdauer eingespeist, um zu untersuchen, welche Variablen bedeutsam zwischen den beiden Gruppen trennen können. Von den eingespeisten Merkmalen trennten vor allem die Items im Helping Alliance Questionnaire (HAQ, Bassler et al., 1995) und die Skala „Durchlässigkeit" im Gießen-Test signifikant zwischen den beiden Patientengruppen. Anhand der Diskriminanzfunktionswerte konnten 73,7% der Patienten als richtig klassifiziert werden. Dieses Ergebnis legt die Vermutung nahe, dass auch die therapeutische Beziehung eine Einflussgröße auf die Behandlungsdauer darstellt.

Es wurden zunächst empirische Kriterien der krankheitsbezogenen Dimension aus nationalen Untersuchungen dargestellt, die einen Zusammenhang zur Behandlungsdauer aufwiesen. Im Folgenden sollen empirische Ergebnisse nationaler Studien zu soziodemographischen Merkmalen und zur Dimension des Therapieverlaufs dargelegt werden.

### 4.2 Zusammenhang zwischen soziodemographischen Merkmalen sowie Merkmalen des Therapieverlaufs und dem Ressourcenverbrauch

Auch **soziodemographische** Merkmale (wie z.B. das Alter oder das Geschlecht) und Merkmale des **Therapieverlaufs** (wie z.B. die therapeutische Beziehung oder Krisen während der psychotherapeutischen Behandlung) können substantielle Zusammenhänge mit dem Verbrauch von Ressourcen in psychotherapeutischer stationärer Behandlung ausüben. Da diese Merkmale aber im Gegensatz zu Variablen der krankheitsbezogenen Dimension oft nur einen geringen Prozentsatz an Varianz (ca. bis 10%) (vergleiche Befunde aus internationalen Studien im Kapitel 3) des Ressourcenverbrauchs oder der Behandlungsdauer aufklären, sollen sie hier lediglich komprimiert dargestellt werden.

Empirische Befunde nationaler Studien zu **soziodemographischen Merkmalen** sind heterogen. Während einige Studien einen Zusammenhang zwischen dem **Geschlecht** eines Patienten und dem Ressourcenverbrauch fanden (Kluge et al., 2002; Roick et al., 2002b; von Heymann et al., 2003), konnten andere Studien diesen Zusammenhang nicht nachweisen (Borgart & Meermann, 1999; Krautgartner et al., 2002; Spitzer et al., 2002). Aber auch in zwei Studien, die einen signifikanten Zusammenhang zwischen Geschlecht und Behandlungsdauer aufzeigen konnten, bleiben die Er-

gebnisse widersprüchlich. Kluge et al. (2002) und von Heymann et al. (2003) berichteten, dass Frauen signifikant länger als Männer behandelt werden. Im Literaturreview von Roick et al. (2002b) fand sich dagegen in sieben Studien eine signifikante Dominanz von Männern in der Gruppe der heavy user, in fünf Studien eine signifikante Dominanz von Frauen. Ebenso heterogen verteilen sich die empirischen Befunde in Bezug auf den **Schul-** oder den **Berufsabschluss** von Patienten. Während Borgart und Meermann (1999) keinen signifikanten Zusammenhang zwischen Schulabschluss und Behandlungsdauer fanden, zeigte die Studie von Richter (1999) signifikante Unterschiede in der Berufsausbildung auf: Patienten, die in die Kategorie A1 der Psychiatrie-Personalverordnung (Kunze & Kaltenbach, 1996) eingestuft worden waren, wiesen zu 67,3% eine abgeschlossene Berufsausbildung auf, Patienten in der Kategorie A2 dagegen zu 47,1%. Ebenso fanden von Heymann et al. (2003) hinsichtlich der zuletzt ausgeübten beruflichen Tätigkeit, dass die bis zum Zeitpunkt der stationären Behandlung noch nicht erwerbstätigen Patienten die längsten Verweildauern aufwiesen.

Weitaus konsistenter sind die empirischen Befunde in Bezug auf den **Familienstand** und das **Alter** der Patienten. So konnten Mestel et al. (2002) und von Heymann et al. (2003) einen signifikanten Zusammenhang zur Behandlungsdauer nachweisen. In den beiden Studien wiesen verheiratete Patienten kürzere Behandlungszeiten auf als ledige oder getrennt lebende Patienten. Auch in der Studie von Borgart und Meermann (1999) wurden ledige Patienten länger behandelt und unterschieden sich signifikant von verheirateten, in Trennung lebenden oder geschiedenen Patienten. Hinsichtlich des Alters zeigte sich, dass jüngere Patienten länger in stationärer Behandlung verweilten (Zielke et al., 1997; Borgart & Meermann, 1999; Krautgartner et al., 2002; Mestel et al., 2002).

Ein weiterer Einflussfaktor auf die Behandlungsdauer ist die **therapeutische Beziehung** oder der therapeutische Verlauf während des stationären Aufenthaltes. Dabei spielen insbesondere Faktoren, wie z.B. durch plötzlichen Arbeitsplatzverlust, Scheidung oder Tod eines Partners verursachte Krisen eine Rolle. So fanden Bassler et al. (1995), dass Patienten mit längerer Behandlungsdauer ihre therapeutische Beziehung in der 1. und 6. Woche der stationären Psychotherapie signifikant positiver einschätzten als Patienten mit kürzeren Verweildauern. In einer 1-Jahres-Katamnese zeigte sich dieser Unterschied nicht mehr. Des Weiteren zeigte sich in der Studie, dass umso länger behandelt wurde, je größer die Schwierigkeiten in der Therapieführung - therapeutenseitig eingeschätzt - waren (z.B. Neigung zum Agieren, eingeschränkte Introspektionsfähigkeit).

In der aktuellen DRG-Diskussion wird wiederholt auf die Bedeutung gleichzeitiger qualitätssichernder Maßnahmen hingewiesen. Zwei Studien, die nun im Folgenden vorgestellt werden sollen, beschäftigten sich mit dem Zusammenhang zwischen Behandlungsdauer und **Therapieerfolg** in der Versorgungspraxis. So konnten Borgart und Meermann (1999) aufzeigen, dass die Behandlungsdauer mit dem Ausmaß der Besserung anstieg und Patienten, die unverändert blieben, die kürzesten Behandlungszeiten hatten. Kritisch merken die Autoren allerdings an, dass es sich dabei um von Therapeutenseite eingeschätzte Besserungen handelte und diese eventuell im Sinne einer kognitiven Dissonanzreduktion diejenigen Patienten, die länger in Behandlung verblieben, auch in ihrem Besserungsgrad höher einschätzten. Bassler et al. (1995) fanden demgegenüber eine signifikante Abnahme der Angstwerte im STAI zu Therapieende und zum 1-Jahres-Katamnesezeitpunkt lediglich für die kürzer behandelte Patientengruppe, nicht aber für die länger behandelte Gruppe. Dieses Muster zeigte sich auch in der Skala „Grundstimmung" des Gießen-Tests, der signifikante Unterschied bestand aber nicht mehr zum Katamnesezeitpunkt. In der Skala „Allgemeiner Beschwerdedruck" des GBB nahmen die Werte zum Ende der Behandlung für beide Gruppen signifikant ab und unterschieden sich erst zum Katamnesezeitpunkt signifikant (länger behandelte Patienten höhere Werte in der Katamnese). Sollte eine DRG-Einführung im stationären Versorgungsbereich von Patienten mit psychischen Störungen zukünftig stärker diskutiert werden, so bedarf es weiterer Studien zum Zusammenhang zwischen Outcome und Behandlungsdauer. In diesem Kontext sollte weiter beforscht werden, inwieweit Therapieerfolg und Verweildauer miteinander interagieren und sich unter bestimmten Bedingungen (z.B. Implementierung eines Klassifikationssystems) verändern.

### 4.3 Fazit

**Zusammenfassend** lässt sich festhalten, dass die vorliegenden empirischen Befunde aus nationalen Studien wichtige Hinweise auf weitere, den Ressourcenverbrauch substantiell mitbestimmende Kriterien liefern. Dabei müssen die verschiedenen Versorgungsbereiche (Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie; Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie; Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen) differenziert betrachtet werden.

Die empirischen Befunde aus der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie sind hinsichtlich des **Kriteriums der Diagnose** nur sehr begrenzt auf den psychosomatisch-

psychotherapeutischen Versorgungsstrang übertragbar: Diejenigen Patienten, die die längsten Verweildauern aufweisen, haben auch signifikant häufiger eine „Schizophrenie“ als Hauptdiagnose, die im psychosomatischen/psychotherapeutischen Versorgungsbereich nur in geringerem Umfang behandelt wird.

Wie bereits dargelegt, stellt die Diagnose allein kein Kriterium dar, welches den Ressourcenverbrauch von Patienten mit psychischen Störungen in stationärer Behandlung adäquat erklären könnte. Aus den empirischen Befunden der nationalen, aber auch der internationalen Studien wird ersichtlich, dass die **Dimension des Schweregrades**, auch in Kombination mit der Hauptdiagnose, einen zentralen Zugang zur angemessenen Abbildung der Ressourcen in stationärer Behandlung von Patienten mit psychischen/psychosomatischen Störungen liefert. Dabei fällt insbesondere ein in Großbritannien entwickeltes Instrument, die Health of the Nation Outcome Scales (HoNOS), auf, welches in zahlreichen internationalen Klassifikationssystemen Verwendung findet. Lenkt man nun den Fokus wieder auf die Ebene der nationalen empirischen Befunde, so lässt sich zusammenfassen, dass der Schweregrad in den einzelnen Studien bislang sehr unterschiedlich operationalisiert wurde. Generell lässt sich aber festhalten, dass je schwerer gestört ein Patient ist, desto länger auch seine Verweildauer ist und umgekehrt.

An dieser Stelle sind noch verschiedene **allgemeine Punkte** anzumerken, die sich einschränkend auf eine Übertragbarkeit der gerade dargestellten Befunde auf die vorliegende Untersuchung auswirken. In den vorgestellten Studien wurden meist sehr spezielle Teilfragestellungen bearbeitet, die von der eigentlichen Konstruktion eines Klassifikationssystems und der Ableitung bzw. Operationalisierung relevanter Dimensionen eines solchen Systems weit entfernt sind. So wurden z.B. Patienten mit einem bestimmten Behandlungskonzept (verhaltenstherapeutisch, Zielke, 1999) untersucht oder nur bestimmte Hauptbehandlungsgruppen berücksichtigt (Kächele für die Studiengruppe MZ-ESS, 1999; Borgart & Meermann, 1999). Des Weiteren griffen viele der nationalen Studien vorwiegend auf Routinedaten zurück, die aufgrund ihrer fehlenden theoretischen wie auch expertengestützten Grundlagen nur in geringem Ausmaß Varianz des Ressourcenverbrauches vorherzusagen vermochten (Borgart & Meermann, 1999; Zielke, 1999; Roick, 2002b; Spießl, 2002).

**Zusammengefasst** wird deutlich, dass es sich beim Schweregrad um eine substanziell bedeutsame Dimension bei der Aufklärung von Ressourcen in stationärer Behandlung handelt. In wie weit nun die im deutschen Sprachraum etablierten Instrumente zur Erhebung des Schweregrades geeignet sind, dieses Merkmal in einer Art und Weise differenziert zu erfassen, damit ein hohes Ausmaß

(erwartet wird ca. 20% - 40% auf der Basis der internationalen Literatur, siehe dazu ausführlich Kap. 3) an Varianzaufklärung erreicht werden kann, soll im folgenden Kapitel überprüft werden.

# 5 Der Schweregrad als ein Prädiktor des Ressourcenverbrauches bei Patienten mit psychischen Störungen

Die differenzierte Bestimmung des **Schweregrades** einer psychischen Erkrankung nimmt wie bereits in Kapitel 3 und 4 dargelegt, einen zentralen Stellenwert bei der Fallgruppenbildung ein. Im folgenden Kapitel soll nun überprüft werden, welche bereits etablierten Instrumente aus dem **deutschsprachigen Raum** zur differenzierten Erfassung des Schweregrades als einer der bedeutsamen Dimensionen bei der Fallgruppenbildung in Betracht gezogen werden können.

## 5.1 Ansätze zur Schweregraderfassung

Erste Ansätze zur Schweregraderfassung gehen bis zum Beginn des 19. Jahrhunderts zurück. Innerhalb der romantischen Psychiatrie bestanden bereits Tendenzen den psychopathologischen Befund im Sinne einer Erfassung der Schwere einer psychischen Störung in einen lebensgeschichtlichen Kontext zu betten und zu erfassen (Hoff, 1997). Im Laufe der Zeit entwickelten sich vor allem zwei verschiedene Zugänge den Schweregrad eines Patienten mit einer psychischen Störung zu erheben: 1. mittels Selbsteinschätzung durch den Patienten oder 2. mittels Fremdeinschätzung durch eine bedeutsame andere Person, oft der Behandler des Patienten.

Bereits im deutschen Sprachraum etablierte **Selbstbeurteilungsinstrumente** sind z.B. die Symptom-Check-List (Franke, 1995) unter Erfassung von eher störungsübergreifenden Aspekten oder das Beck-Depressions-Inventar (Hautzinger et al., 1995) unter Erhebung störungsspezifischer Aspekte. Häufig verwendete **Fremdbeurteilungsinstrumente** in der Bestimmung des psychischen Schweregrades stellen z.B. der Beeinträchtigungs-Schwere-Score (BSS) von Schepank (1995) oder die Skala zur Globalen Erfassung des Funktionsniveaus (GAF, Saß et al., 1998) dar.

### *5.1.1 Selbst- vs. Fremdbeurteilung: Vor- und Nachteile*

**Selbstbeurteilungsverfahren** werden in der klinischen Praxis und Forschung aufgrund verschiedener Vorteile häufiger eingesetzt als Fremdbeurteilungsverfahren (Stieglitz & Freyberger, 2001). So ist ein Großteil klinisch relevanter Informationen nur durch Selbstbeurteilungsverfahren zu erheben (z.B. die Stimmung oder Befindlichkeit eines Patienten). Sie

zeichnen sich ebenfalls dadurch aus, dass sie zeitökonomisch und in sehr kurzen Zeitabständen einsetzbar sind. Die meisten gängigen Verfahren sind teststatistisch überprüft, verfügen partiell aber über Schwächen in ihren psychometrischen Gütekriterien (z.B. keine ausreichenden Angaben zur Reliabilität, meist unbefriedigende Angaben zur Validität oder Fehlen von Normwerten). So besteht z.B. bei der Symptom-Check-List (SCL-90R, Franke, 1995) das Problem, dass sich die Faktorenstruktur nicht eindeutig replizieren lässt (Harfst et al., 2001). Neben Mängeln in teststatistischen Kriterien bedürfen Selbsteinschätzungsinstrumente auch bestimmter anderer Voraussetzungen: Der Patient sollte z.B. compliant sein, über ein Mindestwert an Intelligenz verfügen und eher leicht bis mittel durch seine Symptomatik beeinträchtigt sein. Meist ist diese Verfahrensgruppe nur da von Vorteil, wo der Störungsbereich dem Patienten auch subjektiv zugänglich ist (Stieglitz & Freyberger, 2001).

Diese Defizite von Selbstbeurteilungsverfahren können **Fremdeinschätzungsinstrumente** ausgleichen. Sie sind vorteilhaft, wenn aufgrund der Komplexität der Situation der Patient selbst keine Auskunft geben könnte. Zudem liegt ihr Wert auch sicherlich darin begründet, dass sie eine andere Perspektive der Situation ermöglichen und damit zum zusätzlichen Informationsgewinn beitragen können (Stieglitz et al., 2001). Teststatistisch ist es möglich die Güte der Fremdeinschätzung über die Bestimmung der Interrater-Reliabilität zu prüfen und falls nötig durch Trainings zu verbessern. Empirische Befunde belegen außerdem, dass Fremdbeurteilungen theoretische Konstrukte (wie z.B. Depressivität und Angst) differenzierter abbilden können, als das bei Selbsteinschätzungen möglich wäre (Stieglitz et al., 2001). Stieglitz et al. (2001) kommen aufgrund ihrer Analysen zu dem Schluss, dass sich hinsichtlich der Höhe des Zusammenhanges zwischen Selbst- und Fremdeinschätzung in allen gesichteten Studien große Unterschiede ergeben, aber die Übereinstimmung im Verlauf der Therapie zunimmt. Trotz ihrer Vorteile weisen auch Fremdbeurteilungsverfahren einige Nachteile auf: So sind sie z.B. zeitaufwendiger, da sie oft mit Trainings verbunden sind und können mit Erinnerungs- und Gedächtnisfehlern während der Einschätzung einhergehen.

Stieglitz et al. (2001) kommen zu dem Ergebnis, dass beiden Verfahren **komplementäre Funktion** zukommt, da nicht alle Bereiche gleichermaßen für beide Methoden geeignet sind.

Im Folgenden sollen bereits im deutschen Sprachraum etablierte Selbst- und Fremdeinschätzungsverfahren vorgestellt und abschließend auf eine Eignung zum Einsatz innerhalb eines Fallgruppensystems überprüft werden.

### 5.1.2 *Selbsteinschätzungsinstrumente zur Erfassung des Schweregrades*

Mittlerweile existieren vielfältige Erfassungsmöglichkeiten des Schweregrades einer psychischen Störung durch den Betroffenen selbst. Stieglitz (1999) unterscheidet drei verschiedene Ebenen, auf denen der Schweregrad einer Störung erfassbar ist: **Symptom-, Syndrom- und Diagnoseebene**. Die Symptomebene ist die kleinste Entität, der beobachtbare oder von Patienten berichtete Verhaltensweisen zugrunde liegen. Nach Stieglitz (1999) sind z.B. der Symptomebene die klassischen Arbeiten zur Psychopathologie zuzuordnen, wozu u.a. das mittlerweile vielfach überprüfte System der Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie AMDP zählt (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMDP), 1996). Auf der nächsten Ebene erfolgt dann durch Aufsummierung der einzelnen vorliegenden (beobachteten oder berichteten) Symptome die Syndrombildung. Instrumente, die unmittelbar die Syndromebene tangieren, sind z.B. die Hamilton-Depressions-Skala (Hamilton, 1960) oder das Beck-Depressions-Inventar (Hautzinger et al., 1995). Abschließend erfolgt die Verdichtung der Informationen zur Bildung der Diagnose (Stieglitz, 1999).

**Selbstbeurteilungsinstrumente** zur Erfassung der Beeinträchtigung oder der Schwere einer psychischen Störung sind vor allem auf den **ersten beiden Ebenen** angesiedelt. Nachfolgend soll überprüft werden, in wie weit *bereits im deutschsprachrigen Raum etablierte* Selbsteinschätzungsinstrumente zur Erfassung der Schwere einer psychischen Störung für eine Bildung von Fallgruppen in Frage kommen. Tabelle 6 vermittelt einen Eindruck über *zentrale* Selbsteinschätzungsinstrumente zur Erfassung der Schwere einer psychischen Störung im deutschsprachigen Raum (vgl. Kapitel 4, Abschnitt 4.1.2).

Tabelle 6: Überblick über zentrale Selbsteinschätzungsinstrumente zur Erfassung des Schweregrades einer psychischen Störung

| *Verfahren (Autoren)* | *Zielsetzung* | *Aufbau* | *Praktikabilität* | *Reliabilität* | *Validität* |
|---|---|---|---|---|---|
| **S**ymptom-**C**heck-**L**iste (**SCL-90-R**, (Derogatis, 1977; Franke, 1995) | Erfassung der Psychopathologie | 90 Items<br>9 Skalen (z.B. Depressivität, Phobie)<br>3 Gesamtwerte | Durchführung ca. 10 bis 15 Minuten | Gut | Gut bis auf Schwächen in der faktoriellen Validität |
| **G**iessener **B**eschwerde**b**ogen (**GBB**, Brähler & Scheer, 1983) | Erfassung subjektiv wahrgenommener körperlicher Beschwerden | 57 Items<br>4 Skalen (z.B. Erschöpfung, Magenbeschwerden)<br>Gesamtwert | Durchführung ca. 10 Minuten | Gut | Gut |
| **Be**findlichkeits**s**kala (**Bf-S**, von Zerssen, 2002) | Erfassung subjektiver Gestimmtheit | 28 Adjektiv-Gegensatzpaare<br>Gesamtwert | k.A. | Gut | Gut |
| **B**eschwerden-**L**iste (**BL** von Zerssen, 1976) | Erfassung subjektiver Beeinträchtigung durch körperliche Beschwerden | 24 Items<br>Gesamtwert | k.A. | Gut | Gut |
| **S**hort-**F**orm-**8** (**SF-8,** Ware et al., 2000) | Erfassung gesundheitsbezogener Lebensqualität | 8 Items<br>2 Summenskalen | k.A. | Ausreichend bis Gut | Gut |
| **I**nventar **I**nterpersoneller **P**robleme (**IIP**, Horowitz et al., 1994) | Erfassung interpersoneller Probleme | 64 Items<br>8 Skalen | Durchführung ca. 10 Minuten | Gut | Gut |

**Anmerkungen:** Bewertung der Gütekriterien erfolgte analog zu Stieglitz und Freyberger (2001), Gut = ausreichend empirische Studien durchgeführt mit als zufrieden stellend zu bewertenden Ergebnissen

1977 entwickelte Derogatis an einer Stichprobe von N = 1002 ambulanten Psychiatriepatienten die **Symptom-Check-List-90-R (SCL-90-R)**, welche von Franke (1995) in die deutsche Sprache übersetzt wurde. Das Instrument besteht aus 90 Items mit einem fünf-stufigen Antwortformat („überhaupt nicht“ bis „sehr stark“), die zu 9 Skalen zusammengefasst werden können: *Somatisierung*, *Zwanghaftigkeit*, *Unsicherheit im Sozialkontakt*, *Depressivität*, *Ängstlichkeit, Phobische Angst, Paranoides Denken* und *Psychotizismus* (Franke, 1995). Diese 9 Skalen können wiederum zu 3 Globalwerten verdichtet werden: *Global Severity Index (GSI)*, *Positive Symptom Distress Index (PSDI)* und *Positive Symptom Total (PST)*. Von den drei aufgeführten Maßen stellt der GSI den am häufigsten berichteten Wert dar, welcher psychische Belastung erfasst. Die durchschnittliche Bearbeitungszeit der SCL-90-R dauert ca. 10 bis 15 Minuten. Zur Reliabilität werden im Manual interne Konsistenzen angegeben, die für die Stichprobe der Psychotherapiepatienten in einem guten bis sehr guten Bereich liegen (Cron-

bach´s alpha zwischen .74 und .88). Faktorenanalytische Überprüfungen der SCL-90-R zeigen ein heterogenes Bild: Einige empirische Studien konnten die 9-Faktorenstruktur replizieren, andere dagegen nicht (Franke, 1995). Auch die Interkorrelation der einzelnen Skalen ist hoch. Mittlerweile existiert eine Vielzahl an Validitätsstudien, die eine gute Validität der SCL-90-R belegen (Franke, 1995). Ein Hauptkritikpunkt an der SCL-90-R ist aber neben der zunehmenden Tendenz, praktikable Instrumente einzusetzen, vor allem die faktorielle Struktur des Instrumentes (Harfst et al., 2002). Daher bestehen zurzeit vielfältigste Bemühungen, die SCL in ihrer Itemanzahl zu verkürzen (vgl. hierzu u.a. Harfst et al., 2002).

Der **Gießener Beschwerdebogen (GBB)** ist ein Instrument mit dem das Ziel verfolgt wird, subjektiv wahrgenommene körperliche Beschwerden und des durch die Beschwerden verursachten Leidensdrucks zu erfassen (Brähler & Scheer, 2002). Entwickelt wurde das Instrument auf der Basis einer Beschwerdenliste. Von dieser Liste wurden dann 57 Items ausgewählt. Die ersten Befunde zur faktoriellen Validität wurden bereits 1971 von Zenz veröffentlicht. Anhand einer 5-stufigen Skala (0 = nicht bis 4 = stark) sollen die Patienten angeben, wie stark sie unter den angegebenen Beschwerden leiden. Bei der Auswertung können die Items zu 4 Skalen zusammengefasst werden: *Erschöpfung*, *Magenbeschwerden*, *Gliederschmerzen* und *Herzbeschwerden*. Es kann ein Gesamtbeschwerdedruck, der alle 24 Items umfasst, ermittelt werden. Die Bearbeitungszeit des Fragebogens beträgt ca. 10 Minuten und die Auswertung nimmt ca. 5 Minuten in Anspruch. Inzwischen wurde der GBB (Lang- und Kurzversion) an großen repräsentativen Stichproben eingesetzt. Zur Reliabilität liegen gute Kennwerte vor: So ergeben sich z.B. für die interne Konsistenz des Fragebogens für alle Skalen Werte zwischen Cronbach´s Alpha von .74 bis .91. Die Faktorenstruktur des Instrumentes wurde anhand einer Hauptachsen-Faktorenanalyse mit Varimax-Rotation gewonnen. Zur Validität lässt sich feststellen, dass die einzelnen Skalen des GBB zwischen organischen Befunden in einzelnen klinischen Gruppen (z.B. Herzkatheter- oder Rheumapatienten) weitestgehend differenzieren können. Des Weiteren zeigte sich, dass die Mittelwerte in den Skalen des GBB in einer klinischen Stichprobe signifikant höher waren als in einer Eichstichprobe. Für den GBB liegen auch Normierungsdaten vor.

Originär wurde die **Befindlichkeitsskala (Bf-S)**, ein aus Paaren gegensätzlicher Adjektive bestehender Fragebogen, entwickelt, um zyklothyme Schwankungen bei Patienten mit psychischen Störungen zu erheben (von Zerssen, 2002). Bereits nach den ersten Ergebnissen wurde deutlich, dass wohl eher die Gestimmtheit von Patienten mit der Bf-S erfasst wird. Zur

Konstruktion lässt sich festhalten, dass die Items des Bf-S aus Lehrbuchdarstellungen affektiver Störungen, aus Krankengeschichten und Explorationsprotokollen gewonnen wurden. Zu allen so extrahierten Wörtern wurde im Anschluss ein gegensätzliches Adjektiv hinzugefügt. Es entstanden zwei Parallelformen (Bf-S und Bf-S`), die über jeweils 28 Gegensatzpaare an Adjektiven verfügen z.B. frisch-matt, froh-schwermütig oder lustig-weinerlich. Der Patient gibt bei der Beantwortung dieser Adjektivpaare an, welche der beiden Eigenschaften eher auf ihn zutreffen. Bei der Testauswertung erhält man einen Gesamtwert als Summe über alle vergebenen Punkte für die Antworten (2 Punkte für ein Kreuz bei einer depressiv bezeichnenden Eigenschaft, 0 Punkte bei der gegenteiligen Antwort und 1 Punkt bei einer „weder-noch“ Antwort). An einer bevölkerungsrepräsentativen Stichprobe von N = 1.761 Erwachsenen wurden die psychometrischen Eigenschaften der beiden Parallelformen überprüft. So kann festgehalten werden, dass die Objektivität gegeben ist (standardisiertes Verfahren) und die internen Konsistenzen für beide Parallelformen im zufrieden stellenden Bereich angesiedelt sind (Cronbach´s Alpha über .90). Die faktorielle Validität kann als gut betrachtet werden: In einer unabhängigen Hauptkomponentenanalyse konnte die einfaktorielle Struktur des Fragebogens repliziert werden. Auch die interne und externe Validität der Bf-S kann als gegeben bewertet werden: So korreliert die Skala sowohl mit anderen Skalen zur Erfassung von Befindlichkeit in einem mittleren bis hohen Bereich, sie kann zwischen gesunden und kranken Kontrollpersonen unterscheiden und weist Änderungssensitivität auf.

Von Zerssen entwickelte 1976 die **Beschwerden-Liste (B-L)**, sowie eine Parallelform (B-L´), welche die subjektive Beeinträchtigung durch körperliche Beschwerden und des Allgemeinbefindens erfasst. Auf der Basis von Item- und Faktorenanalysen wurden die Items des Fragebogens aus einer Liste von Beschwerden, die häufig von körperlich oder psychisch Erkankten angegeben wurden, extrahiert. Die B-L und auch die B-L´ bestehen aus 24 Substantiven (z.B. Schwächegefühl, Reizbarkeit oder Zittern) die 4-fach gestuft zu beantworten sind (von „stark“ bis „gar nicht“). Zur Auswertung steht sowohl ein Summenscore als auch ein globales Beschwerdemaß, errechnet aus dem Mittelwert der Summenscores der beiden Parallelformen, zur Verfügung. Die beiden Fragebögen wurden anhand einer bevölkerungsrepräsentativen Erhebung Erwachsener hinsichtlich ihrer psychometrischen Eigenschaften überprüft und normiert. Auch hier ist die Objektivität des Verfahrens gegeben und die interne Konsistenz der beiden Fragebögen kann als zufrieden stellend bewertet werden. Die Paralleltest-Reliabilität liegt zwischen .85 und .88 und kann als gut betrachtet werden. Zur faktoriellen

Validität lässt sich aussagen, dass die gefundene Struktur eines Generalfaktors durch unabhängige Analysen repliziert werden konnte. Ergebnisse zur Korrelation mit anderen Skalen oder nur geringe Korrelation mit z.B. einem Untertest des HAWIE (Hamburg Wechsler Intelligenztest) sprechen für eine gute konvergente und divergente Validität der B-L und B-L´. Eine Vielzahl von Studien (vgl. hierzu von Zerssen, 2002) belegen die Änderungssensitivät des Verfahrens.

Der **Short-Form 8 (SF-8)** ist ein generisches Instrument zur Erfassung des Gesundheitsstatutes eines Patienten (Ware et al., 2000). Ziel bei der Konstruktion der SF-8 war es, eine Kurzform des SF-36 zu entwickeln, in der 1 Item für die 8 Gesundheitskonzepte der SF-36 steht und welches in größeren Studien praktikabel einsetzbar sein sollte. Obwohl sich die SF-8 zum Teil mit der SF-36 inhaltlich überschneidet, sind keine der Items der SF-8 mit der Langversion identisch. Mittlerweile handelt es sich genauso wie bei der SF-36 um ein international weit verbreitetes und in viele Sprachen übersetztes Instrument. Das Antwortformat der SF-8 variiert von 5- bis 6-stufiger Skala. Sowohl mit der SF-8 als auch mit dem SF-36 lassen sich zwei Summenskalen („Psychische und Körperliche Summenskala“) bilden. Zur Reliabilität lässt sich festhalten, dass interne Konsistenzen in einem zufrieden stellenden Bereich zwischen .70 und .88 variieren. Test-Retest Reliabilitäten (Zeitraum: 4 Wochen) befinden sich mit Werten zwischen .59 und .70 in einem guten Bereich. Eine Überprüfung der Validität wurde in insgesamt 14 Studien durchgeführt: Konvergente Validität ist gegeben (die Skalen der SF-8 korrelieren signifikant mit den Skalen des SF-36) und unabhängige Faktorenanalysen anhand großer repräsentativer amerikanischer Stichproben konnten die 2-Faktorenlösung der SF-8 replizieren. In klinischen Studien (Personen ohne bzw. mit körperlichen oder psychischen Problemen) konnte die diskriminative Valididät der SF-8 gut belegt werden: Die körperliche Summenskala wies in der Gruppe der Personen mit körperlichen Problemen die höchsten Werte auf, die Psychische Summenskala dagegen in der Gruppe der Personen mit psychischen Problemen. Auch für die SF-8 ist eine Normierung im Rahmen einer amerikanischen Gesundheitsuntersuchung durchgeführt wurden.

Einen völlig anderen Bereich als die oben beschriebenen Instrumente erhebt das **Inventar zur Erfassung Interpersoneller Probleme (IIP,** Horowitz et al., 1994). Basierend auf Beobachtungen mit Psychotherapiepatienten, dass diese vor allem in den Therapiesitzungen Probleme im interpersonellen Bereich berichteten, entwickelten Horowitz et al. 1988 das IIP. Auf der Basis von Psychotherapiegesprächen sammelten und kategorisierten die Autoren zunächst 129

Items. Von Alden et al. (1990) wurde die Skalenstruktur des Fragebogens entwickelt (die sogenannte Circumplex-Struktur), die 8 Skalen der 64 Items umfasst: *Zu autokratisch/dominant, zu streitsüchtig/konkurrierend, zu abweisend/kalt, zu introvertiert/sozial vermeidend, zu selbstunsicher/unterwürfig, zu ausnutzbar/nachgiebig, zu fürsorglich/freundlich* und *zu expressiv/aufdringlich*. Eine Übersetzung des Instrumentes in die deutsche Sprache erfolgte von Horowitz, Strauß und Kordy im Jahr 2000. Die durchschnittliche Durchführungszeit wird von den Autoren für die 64-Item umfassende Version des IIP mit 10 bis 15 Minuten angegeben. Zu den psychometrischen Eigenschaften des IIP lässt sich festhalten, dass die Objektivität durch die Standardisierung gegeben ist. Hinsichtlich der Reliabilität ergeben sich eher moderate interne Konsistenzen der acht Skalen: Die Werte schwanken zwischen .36 und .64. In unterschiedlichen Stichproben konnte die Circumplex-Struktur des Fragebogens repliziert werden. Auch die externe Validität des IIP ist gegeben: So existieren eine Vielzahl an Studien, die einen Zusammenhang zwischen externen Faktoren (z.B. andere Persönlichkeitsmerkmale) und den Skalen des IIP herstellen können. Auch hinsichtlich der prognostischen Validität hat sich das Instrument bewährt. Es zeigte sich, dass ein höheres Maß an interpersonell berichteten Problemen in Zusammenhang zum Outcome von Patienten stand. Zur Änderungssensitivität lässt sich festhalten, dass Werte im IIP nach einer Therapie sich nur wenig ändern und sich deutlichere Veränderungen in Katamnesen zeigen. Der Fragebogen wurde auch normiert. Mittlerweile existieren Bemühungen eine Kurzversion des IIPs für einen praktikablen Einsatz in größeren Studien oder für Routineerhebungen zu entwickeln (Harfst et al., 2004b).

### *5.1.3 Fremdeinschätzungsinstrumente zur Erfassung des Schweregrades*

Neben den Selbsteinschätzungsinstrumenten kommt gerade bei der Erfassung des Schweregrades **Fremdeinschätzungsinstrumenten** eine große Bedeutung zu: Stellen sie doch meist die Symptomatik des Patienten aus einem anderen Blickwinkel dar. Unter zunehmender Belastung derjenigen, die die Beurteilungen vornehmen (z.B. Therapeuten), gehen auch hier die Bemühungen dahin, Instrumente zu verwenden, die möglichst praktikabel und zuverlässig das Wesentliche erfassen. Tabelle 7 stellt auch hier wieder eine Auswahl an *gut etablierten Fremdeinschätzungsinstrumenten* bereit, die nachfolgend vorgestellt und abschließend unter der Prämisse psychometrischer Eigenschaften bewertet werden sollen.

Tabelle 7: Überblick über zentrale Fremdeinschätzungsinstrumente zur Erfassung des Schweregrades einer psychischen Störung

| ***Verfahren (Autoren)*** | ***Zielsetzung*** | ***Aufbau*** | ***Praktikabilität*** | ***Reliabilität*** | ***Validität*** |
|---|---|---|---|---|---|
| AMDP-System (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMDP), 1996) | Erfassung psychopathologischer Befunde | 4 Dimensionen (Soziodemographie, Anamnese, psychischer und somatischer Befund)<br>140 Items<br>9 Subskalen<br>3 übergeordnete Skalen | Bearbeitungszeit: 30 bis 60 Minuten<br>Training erforderlich | Für 68 Items moderat bis zufriedenstellende Werte | Gut[1] |
| Beeinträchtigungs-Schwere-Score (BSS, Schepank, 1995) | Erfassung der Schwere einer psychischen Erkrankung | 3 Dimensionen<br>Gesamtwert | Bearbeitungszeit: ca. 2 Minuten | Gut | Gut[1] |
| Brief Psychiatric Rating Scale (BPRS Overall & Gorham, 1976) | Erfassung des Schweregrades psychischer Symptome | 18 Symptomkomplexe<br>Gesamtwert<br>5 Subskalen | Bearbeitungszeit: ca. 25 Minuten | Gut | Gut[1] |
| CGI (National Institute of Mental Health, 1976) | Nutzen-Risiko-Bewertung bei medikamentöser Behandlung | 3 Items | k.A. | Moderat | Gut[2] |
| Skala zur Globalen Erfassung des Funktionsniveaus (GAF, Endicott et al., 1976) | Erfassung des Funktionsniveaus | Gesamtwert | Bearbeitungszeit: ca. 1 Minute | Moderat bis gut | Gut[1] bis auf Schwächen in der diskriminierenden Validität |

***Anmerkungen:*** [1]Bewertung der Gütekriterien erfolgte nach Stieglitz und Freyberger (Stieglitz & Freyberger, 2001), Validität Gut = ausreichend empirische Studien durchgeführt mit guten Ergebnissen; [2]Bewertung aufgrund vorliegender empirischer Befunde aus Validitätsstudien zu den einzelnen Instrumenten

Das **AMDP System** wurde mit der Zielsetzung der Dokumentation psychopathologischer Befunde von der Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie entwickelt (Collegium Internationale Psychiatriae Scalarum (CIPS), 1996c). Es handelt sich um ein umfassendes System, welches gleichermaßen soziodemographische Daten, anamnestische Daten, einen psychischen und somatischen Befund des Patienten erhebt. Für die vorliegende Arbeit ist der *psychische Befund* von besonderem Interesse, da er unter Umständen unmittelbar zur Erfassung des Schweregrades herangezogen werden könnte. Charakteristisch für den psychischen Befundteil des AMDP-Systems ist eine vierfach gestufte Beurteilung („nicht vorhanden" bis „schwer ausgeprägt") von **zwölf Merkmalen** vor allem für Patienten mit einer F1 (Alkohol- oder Substanzabusus) oder F2 Störung der ICD-10 (Störungen des schizophrenen Formenkreises): *Bewussteinsstörungen*, *Orientierungsstörungen*, *Aufmerksamkeits- und Gedächtnisstörungen*, *Befürchtungen und Zwänge*, *Wahn*, *Sinnestäuschungen*, *Ich-Störungen*, *Störungen der Affektivität*, *Antriebs- und psychomotorische Störungen*, *Circadiane*

*Besonderheiten* und *Andere Störungen*. Diese zwölf Merkmale sind wiederum in Subitems untergliedert, wie Abbildung 8 exemplarisch für das Item „Ich-Störungen" zeigt. Insgesamt enthält das AMDP-System psychischer Befund 100 Merkmale. Es lassen sich zusammengefasst aus den Symptomen des psychischen und des somatischen Befundes 9 Syndromskalen (z.B. depressives oder manisches Syndrom) sowie 3 übergeordnete Syndrome (paranoid-halluzinatorische Symptomatik, depressive und psychoorganische Symptomatik) bilden. Die zur Beurteilung nötigen Informationen können aus Aufnahmegesprächen durch verschiedene Mitarbeiter des Teams (Arzt, Psychologe, Pflegepersonal) oder auch durch Angehörige des Patienten gewonnen werden. Um eine Beurteilung im AMDP-System vornehmen zu können, sollte der Anwender mit den klinischen Konzepten der Items vertraut sein und über ein Training verfügen. Die Bearbeitungszeit beträgt 30 bis 60 Minuten. Zur Reliabilität lässt sich festhalten, dass vorliegende Interraterkoeffizienten (kappa-Werte) für 68 Items moderate bis zufrieden stellende Werte zwischen .40 und .68 einnehmen, wenn „Symptom vorhanden" oder „nicht vorhanden" eingeschätzt wird; 13 Items liegen unter einem kappa von .40. Bei den verbliebenen Items konnten keine Interraterreliabilitäten ermittelt werden, da diese Items zu selten auftreten. Die Interraterreliabilität für die Syndromskalen liegen mit Werten zwischen .63 und .93 in einem zufrieden stellenden Bereich. Aspekte zur diskriminanten Validität hinsichtlich verschiedener Patientengruppen können als gut bewertet werden: Das AMDP-System kann auf syndromaler Ebene zwischen Krebspatienten und Patienten in stationärer psychotherapeutisch/psychiatrischer Behandlung unterscheiden (Baumann & Stieglitz, 1989). Validitätsnachweise bezüglich des Zusammenhanges mit anderen Fremdbeurteilungsinstrumenten (Brief Psychiatric Rating Scale BPRS, Collegium Internationale Psychiatriae Scalarum (CIPS), 1996a) sind als gut zu bewerten, erwartungsgemäß zeigen sich nur geringe Korrelationen zu Selbsteinschätzungsinstrumenten (Frankfurter Beschwerdebogen FBF und Eigenschaftswörterliste EWL). Ergänzende Arbeiten zu Reliabilitäts- und Validitätsaspekten liefert die Übersichtsarbeit von Baumann und Stieglitz (1989).

| | Nicht vorhanden ▫ **Ich-Störungen** | | | | |
|---|---|---|---|---|---|
| | **Nicht vorhanden** | **Leicht** | **Mittel ausgeprägt** | **Schwer** | **Keine Aussage** |
| **Derealisation** | ▫ | ▫ | ▫ | ▫ | ▫ |
| **Depersonalisation** | ▫ | ▫ | ▫ | ▫ | ▫ |
| **Gedankenausbreitung** | ▫ | ▫ | ▫ | ▫ | ▫ |
| **Gedankenentzug** | ▫ | ▫ | ▫ | ▫ | ▫ |
| **Gedankeneingebung** | ▫ | ▫ | ▫ | ▫ | ▫ |
| **Fremdbeeinflussungserleben** | ▫ | ▫ | ▫ | ▫ | ▫ |

Abbildung 8: Ausschnitt aus dem psychischen Befunderhebungsbogen des AMDP-Systems exemplarisch für das Item „Ich-Störungen“ (Collegium Internationale Psychiatriae Scalarum (CIPS), 1996c)

Die **Brief Psychiatric Rating Scale** (Overall & Gorham, 1976) ist das Produkt der Verkürzung der Mulitdimensional Scale for Rating Psychiatric Patients MSRPP (Lorr et al., 1953) und der Inpatient Multidimensional Psychiatric Scale IMPS (Lorr et al., 1963). Overall und Gorham (1976) bildeten auf der Grundlage von Expertenratings und multivariaten Techniken zunächst 16 später dann 18 Symptomkomplexe, die 7-stufig einzuschätzen sind („nicht vorhanden“ bis „extrem stark“). 1986 wurde die BPRS von Hiller, von Zerssen, Mombour und Wittchen in die deutsche Sprache übersetzt (Hiller et al., 1986). Bei den 18 Symptomkomplexen handelt es sich um folgende Inhalte: *Körperbezogenheit*, *Angst*, *Emotionale Zurückgezogenheit*, *Zerfall der Denkprozesse*, *Schuldgefühle*, *Gespanntheit*, *Maniertheit-Affektiertheit-Positur*, *Größenideen*, *Depressive Stimmung*, *Feindseligkeit*, *Misstrauen-paranoide Inhalte*, *Halluzinationen*, *Motorische Verlangsamung*, *Unkooperatives Verhalten*, *Ungewöhnliche Denkinhalte*, *Affektive Abstumpfung-Verflachung*, *Erregung* und *Orientierungsstörungen* (Collegium Internationale Psychiatriae Scalarum (CIPS), (1996a), siehe zur exemplarischen Veranschaulichung Abbildung 9). Aus dem Rating der Einzelitems lassen sich sowohl ein Gesamtscore (= Ausmaß der psychischen Gestörtheit) als auch fünf Skalenwerte bilden (Angst/Depression, Anergie, Denkstörung, Aktivierung und Feinseligkeit/Misstrauen). An dieser Stelle ist noch aufzuführen, dass die Testkonstrukteure Ärzte oder Psychologen als Beurteiler sowie die Zusammenarbeit von zwei Urteilern empfehlen. Die Informationen für das Rating werden aus einem ca. 20 Minuten dauernden Interview gewonnen, welches durch eine

gezielte Befragung ergänzt wird. Die Bearbeitungsdauer beschränkt sich auf ca. 3 Minuten. Die Reliabilität der BPRS lässt sich größtenteils als gut bewerten: Interraterreliabilitäten für die Faktoren nehmen Werte zwischen .52 und .90 ein, für den Gesamtwert liegen die Werte zwischen .87 und .97. Auch die Konstrukt- und kriterienbezogene Validität der BPRS gilt als gut belegt (Collegium Internationale Psychiatriae Scalarum (CIPS), 1996a): So konnte z.B. zwischen dem BPRS-Gesamtwert und der MSRPP eine Korrelation von .93 ermittelt werden.

| Nicht vorhanden | Sehr gering | Gering | Mäßig | Mäßig stark | Stark | Extrem stark |
|---|---|---|---|---|---|---|
| ① | ② | ③ | ④ | ⑤ | ⑥ | ⑦ |

**Halluzinationen**
Wahrnehmungen ohne entsprechende normale äußere Reize. Bewerten Sie nur solche Erlebnisse, die laut Patient in der letzten Woche aufgetreten sind und die sich – so wie beschrieben werden – deutlich vom Denken und der Vorstellung Normaler abheben

Abbildung 9: Ausschnitt aus der Brief Psychiatric Rating Scale am Beispiel des Items „Halluzinationen" (Collegium Internationale Psychiatriae Scalarum (CIPS), 1996a)

Die **Clinical Global Impressions Scale (CGI)** wurde 1970 vom National Institute of Mental Health im Rahmen der Schizophrenieforschung entwickelt (Collegium Internationale Psychiatriae Scalarum (CIPS), 1996b). Hauptsächliches Ziel des Instrumentes ist die Nutzen-Risiko-Bewertung bei medikamentöser Behandlung von Patienten mit psychischen Störungen. Die Skala besteht aus 3 Items, wobei die ersten beiden Items in 7-stufiger Ausprägung beurteilt werden sollen: Schweregrad der Krankheit, Gesamtbeurteilung der Zustandsänderung und Therapeutische Wirksamkeit/Unerwünschte Wirkungen (Therapie-Risiken). Abbildung 10 veranschaulicht exemplarisch die Einschätzung des Schweregrades mittels der CGI Skala. Die Beurteilung sollte von einem Arzt vorgenommen werden. Eine Besonderheit und Folge einer überarbeiteten Version stellt die Anweisung dar, dass der Beurteiler jeweils den Bezugsrahmen des „typischen Patienten" und nicht seinen gesamten Erfahrungsschatz zugrunde legen sollte. Die Interraterreliabilitäten für die drei Items der CGI Skala sind in einem eher moderaten Bereich (Werte zwischen .35 und .66), Retest-Reliabilitäten nehmen aber zufrieden stellende Werte ein (Werte zwischen .15 und .81). Eine Untersuchung zur Validität der CGI Ska-

la fand heraus, dass bei Demenzpatienten die Beurteilung des Schweregrades der Erkrankung (Item 1, siehe Abbildung 10) wesentlich durch diagnostische Kriterien (DSM-III-R) bestimmt wird.

**1. Schweregrad der Krankheit**
Ziehen Sie Ihren gesamten Erfahrungsschatz an dieser Art von Kranken in Betracht,
und geben Sie an, wie hoch Sie den jetzigen Grad der seelischen Erkrankung des Patienten einschätzen.

⓪ Nicht beurteilbar
① Patient ist überhaupt nicht krank.
② Patient ist ein Grenzfall psychiatrischer Erkrankung.
③ Patient ist nur leicht krank.
④ Patient ist nur mäßig krank.
⑤ Patient ist deutlich krank.
⑥ Patient ist schwer krank.
⑦ Patient gehört zu den extrem schwer Kranken.

Abbildung 10: Ausschnitt aus der Clinical Global Impressions Scale (CGI) für die Beurteilung der Schwere der Erkrankung (Collegium Internationale Psychiatriae Scalarum (CIPS), 1996b)

Zwei sowohl in der Forschung als auch in der klinischen Praxis in Deutschland häufig verwendeten Instrumente zur Fremdeinschätzung des Schweregrades ist der **Beeinträchtigungs-Schwere-Score** (BSS) von Schepank (1995) und die **Skala zur Globalen Erfassung des Funktionsniveaus** (GAF) (Endicott et al., 1976; Saß et al., 1998).

Die **Global Assessment of Functioning Scale** oder **Skala zur Globalen Erfassung des Funktionsniveaus (GAF)** erfasst das allgemeine Funktionsniveau unter Berücksichtigung psychischer, sozialer und beruflicher Aspekte von Patienten mit psychischen Störungen (Saß et al., 1998) und bildet die Achse V des DSM-IV. Zurückgehend auf die Entwicklungen einer Globalskala von Luborsky im Jahre 1962 („Health-Sickness Rating Scale HSRS), konstruierte Endicott die Global Assessment Scale als überarbeitete Skala der HSRS (American Psychiatric Association, 2002). Die Beurteilung des Ausmaßes der Schwere einer Funktionsbeeinträchtigung wird anhand einer Skala von 1-100 getroffen, wobei ein Wert von 100-91 eine

hervorragende Leistungsfähigkeit und ein Wert von 10-1 eine sehr stark beeinträchtigte Leistungsfähigkeit widerspiegelt[8]. Beeinträchtigungen aufgrund körperlicher Erkrankung werden in die Einschätzung nicht miteinbezogen. Obwohl die empirische Basis als noch nicht ausreichend beurteilt werden muss, zeigen doch vorliegende Befunde gute Gütekriterien für die GAF-Skala. Zur Reliabilität der Skala werden Werte zwischen .54 und .90 berichtet, wobei sich diese große Spannweite durch den Trainings- bzw. Erfahrungsstand der Beurteiler erklären lässt. Auch hinsichtlich der Validität können gute Ergebnisse verzeichnet werden: So korreliert die GAF-Skala z.B. mit der WHO-disability-diagnostic-scale mit r = -.64 (American Psychiatric Association, 2002).

Der **Beeinträchtigungsschwerescore** (**BSS,** Schepank, 1995) erhebt wie schwer ein Patient im psychischen, körperlichen und sozialkommunikativen Bereich durch seine Symptomatik beeinträchtigt ist. Die Entwicklung des Scores erfolgte im Rahmen der Beforschung potentieller erblicher Faktoren in der Ätiologie von psychischen Erkrankungen. Auf der Basis von 100 Zwillingspaaren wurde der Vorläufer der BSS „der Neurose-Schwere-Score“ modifiziert und gekürzt (Schepank, 2002). Der BSS wird seit 1975 in der Mannheimer Psychosomatischen Klinik eingesetzt und gehört mittlerweile zur Basisdokumentation in der stationären psychosomatisch-psychotherapeutischen Krankenhausbehandlung. Es steht ein Profilwert zur Verfügung, der die Bevorzugung der Beeinträchtigung wider gibt und ein Summenwert, der die Gesamtbeeinträchtigung wider spiegelt (Schepank, 1995). Laut dem Testautor wird eine gründliche und persönliche Diagnostik (ca. 1 bis 1,5 h Interview) eines in der Beurteilung psychischer Krankheiten erfahrenen Experten für eine Einschätzung im BSS vorausgesetzt (Schepank, 1995). Der Beurteilungszeitraum kann sich zwischen den letzten 7 Tagen oder den letzten 12 Monaten bewegen, im Gegensatz zur GAF-Skala, bei welcher die Beurteilung zum aktuellen Zeitpunkt vorgenommen wird (Lange & Heuft, 2002). Interraterreliabilitäten können als zufrieden stellend bewertet werden mit Werten zwischen .89 und .99, für die Subdi-

---

[8] Beispielhaft sollen für die Werte 100-91 und 90-81 die Operationalisierungen aufgeführt werden. **100-91** = Hervorragende Leistungsfähigkeit in einem breiten Spektrum von Aktivitäten; Schwierigkeiten im Leben scheinen nie außer Kontrolle zu geraten; keine Symptome. **90-8**1 = Keine oder nur minimale Symptome (z.B. leicht Angst vor einer Prüfung), gute Leistungsfähigkeit in allen Gebieten, interessiert und eingebunden in ein breites Spektrum von Aktivitäten, sozial effektiv im Verhalten, im allgemein zufrieden mit dem Leben, übliche Alltagsprobleme oder -sorgen (z.B. nur gelegentlicher Streit mit einem Familienmitglied) (Saß et al., 1998).

mensionen zwischen .68 und .97. Die bislang vorliegenden Studien zur Validität zeigen ebenfalls gute Ergebnisse (Schepank, 2002). Abbildung 11 zeigt die Ratingmaske des BSS.

**1. Körperlicher Beeinträchtigungsgrad (0-4):**

| Gar nicht | Geringfügig | Deutlich | Stark | Extrem |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

**2. Psychischer Beeinträchtigungsgrad (0-4):**

| Gar nicht | Geringfügig | Deutlich | Stark | Extrem |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

**3. Sozialkommunikativer Beeinträchtigungsgrad (0-4):**

| Gar nicht | Geringfügig | Deutlich | Stark | Extrem |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

Abbildung 11: Beurteilungsmaske der drei Dimensionen des Beeinträchtigungsschwerescores (BSS, Schepank, 2002).

**Anmerkungen:** Die Skalenstufen 0 bis 4 sind dabei wie folgt operationalisiert: 0 = praktisch ohne jegliche Störung; 1 = eine Symptommanifestation ist in geringem Ausmaß zwar vorhanden, sie wird aber als leicht eingeschätzt; 2 = die Symptomatik ist unübersehbar vorhanden; 3 = eine ausgeprägte und schon erhebliche Beeinträchtigung des Individuums auf der entsprechenden Skalendimension ist vorhanden und 4 = die Symptomatik ist so stark, dass sie auf der zu beurteilenden Dimension zu einer kaum noch überbietbaren Beeinträchtigung des Individuums im Lebensalltag führt.

In einer Untersuchung verglichen Lange und Heuft (2002) diese beiden Fremdeinschätzungsinstrumente hinsichtlich ihrer teststatistischen Güte und ihres Differenziertheitsgrades anhand einer Stichprobe von 263 Ambulanz- und Konsilpatienten einer Klinik für Psychosomatik und Psychotherapie. Von den Autoren wurde dabei zu beiden Instrumenten angemerkt, dass bislang noch größere Studien zu teststatistischen Gütekriterien fehlen, wobei sich zusätzlich bei der GAF-Skala Validitätsprobleme durch die kombinierte Erfassung von psychischer Symptomatik und sozialem Funktionsniveau abzeichnen. Auch Laireiter et al. (2001) kommen bei der Bewertung der teststatistischen Gütekriterien des BSS zu dem Schluss, dass zwar ausreichend Studien zur Überprüfung der Reliabilität mit guten Ergebnissen vorliegen, aber dennoch weitere Studien zur Prüfung der Validität nötig wären. Bei der Bewertung der GAF-Skala sind sowohl für die Reliabilität als auch für die Validität Hinweise auf ihre Testgüte vorhanden. Laireiter et al. (2001) sprechen sich aber auch hier dafür aus, dass weitere Überprüfungen erforderlich sind. Abschließend plädieren Lange und Heuft (2002) dafür, dass eine Kombination beider Verfahren sinnvoll und notwendig wäre, um das gesamte Spektrum psychischer Störungen bezüglich des Schweregrades und der sozialen Funktionsfähigkeit erfassen zu können. Das würde allerdings zu einem erheblichen Mehraufwand führen, weil dann Trainings für beide Verfahren erforderlich wären.

### 5.2 Abschließende Bewertung zur Eignung der Selbst- und Fremdeinschätzungsinstrumente zum Einsatz innerhalb eines Klassifikationssystems

In diesem Kapitel wurden zentrale, bereits im deutschsprachigen Raum etablierte Selbst- und Fremdeinschätzungsinstrumente mit ihren psychometrischen Eigenschaften vorgestellt. Im Folgenden soll eine abschließende Bewertung zur Eignung eines Einsatzes dieser Instrumente innerhalb der Bildung von Fallgruppen für Patienten mit psychischen Störungen vorgenommen werden.

Wie eingangs erwähnt, besitzen Selbst- und Fremdeinschätzungen auch unter der Perspektive der Entwicklung eines Klassifikationssystems bei Patienten mit psychischen Störungen Vor- und Nachteile (Stieglitz et al., 2001). Hinsichtlich der **Fremdeinschätzung** könnte es von Nachteil sein, dass mit einem Klassifikationssystem Funktionen verknüpft sind, die anreizorientiert sind. So könnten z.B. Behandler den Schweregrad der Erkrankung ihrer Patienten überschätzen, um z.B. in Verhandlungen mit Kostenträgern höhere Fallpauschalen zu erwirken. Bereits Hirdes et al. (2003) beschrieben dieses Verhalten als Manipulierbarkeit oder „Gaining“ von Daten: „For example, variables that cannot be observed by an independent, external rater would be easy to manipulate in the hope of gaining additional funding. The greater the reliance on subjective interpretation by the assessor, the greater the risk that the variable can be systematically overestimated with relative impunity.” (zitiert nach Hirdes et al., 2003, S. 8).

**Selbsteinschätzungen** können allerdings dort an ihre Grenzen stoßen, wo die Betroffenen so schwer erkrankt sind, dass sie selbst keine Auskunft mehr geben können (z.B. ein Patient mit einer akuten schizophrenen Episode). Weiterhin zeigen empirische Befunde auf, dass Selbstbeurteilungen bei Patienten mit starker psychischer Beeinträchtigung häufiger anscheinend allgemeine Klagsamkeit erfassen (Vassend & Skrondal, 1999). Innerhalb der Entwicklung eines Klassifikationssystems für Patienten mit psychischen Störungen bedarf es daher eines Instrumentes, welches auch in Bereichen größerer Symptombelastung gut differenzieren kann, um möglichst viel Varianzaufklärung am Ressourcenverbrauch bzw. an der Behandlungsdauer als ein Kriterium der Güte des Systems erklären zu können.

International komplementär zu DRG-Systemen entwickelte Patientenklassifikationen beruhen auf Fremdeinschätzungen durch Ärzte, Psychologen, Sozialarbeiter und Pflegekräfte (Taube

et al., 1984b; Young et al., 1985; Mitchell et al., 1987; Ashcraft et al., 1989; Horn et al., 1989; Buckingham et al., 1998; Gaines et al., 2003; Hirdes et al., 2003). Da sich das stationäre Versorgungssystem psychisch Kranker in der Bundesrepublik Deutschland vom internationalen Versorgungssystem unterscheidet (vgl. hierzu ausführlich Kapitel 1 „Einleitung und Bedeutsamkeit der Fragestellung) und in dem Versorgungssektor, in dem die vorliegende Untersuchung durchgeführt werden soll, Patienten eher selten in dem Ausmaß beeinträchtigt sind, dass sie keine Auskunft mehr geben könnten, sollten *sowohl Selbst- als auch Fremdeinschätzungsinstrumente* als Komponenten des zu entwickelnden Klassifikationssystems herangezogen werden.

Ein Klassifikationssystem für Patienten mit psychischen Störungen sollte sich aufgrund seiner vielfältigen möglichen Funktionen (z.B. als Instrument zum Krankenhausbetriebsvergleich oder als internes Ressourcenmanagementinstrument) in erster Linie durch seine **Praktikabilität** auszeichnen, um in der Routine angewendet zu werden. Die hier aufgeführten **Selbsteinschätzungsinstrumente** verfügen alle weitest gehend über gute bis sehr gute psychometrische Kennwerte. Große Unterschiede bestehen allerdings hinsichtlich ihrer Praktikabilität. Für die vorliegende Untersuchung können die SCL und die SF als zentrale Selbsteinschätzungsinstrumente zur Erfassung der Schwere einer Erkrankung bewertet werden. Wie bereits erwähnt liegen beide Instrumente bereits in gekürzter Version vor und können damit als praktikabel eingestuft werden. Ein weiteres zentrales Instrument zur Erhebung der Schwere einer Erkrankung im interpersonellen Problembereich ist der IIP, der ebenso in einer gekürzten Version vorliegt. Alle weiteren aufgeführten Instrumente, wie der Gießener Beschwerdebogen, die Befindlichkeitsskala und die Beschwerdenliste liegen noch nicht in gekürzter Version vor und erfassen zudem Bereiche, die durch die eben als zentral aufgeführten Selbsteinschätzungsinstrumente bereits erhoben werden. Sie sollen deswegen in der vorliegenden Untersuchung keine Berücksichtigung finden.

Hinsichtlich der aufgeführten **Fremdeinschätzungsinstrumente** lässt sich zunächst festhalten, dass einige Messverfahren über eher moderate psychometrische Eigenschaften verfügen (z.B. das AMDP-System oder der CGI) und deswegen keine Berücksichtigung finden sollen. Ein weiteres Beurteilungskriterium bei der Auswahl eines geeigneten Instrumentes sollte sich neben der **Praktikabilität** auch auf die Möglichkeit zur **differenzierten Erfassung des Schweregrades** einer psychischen Erkrankung beziehen. Bei der Bewertung der Praktikabilität fällt auf, dass kaum eines der aufgeführten Systeme durch in der Routine realisierbare Be-

arbeitungszeiten gekennzeichnet ist: So z.B. im Fall der BPRS, bei der durchschnittlich 25 Minuten zur Beantwortung benötigt werden. Von den aufgeführten Fremdeinschätzungsinstrumenten kommen deswegen hauptsächlich nur zwei in Betracht: der BSS und der GAF. Es lässt sich allerdings festhalten, dass unabhängig von den psychometrischen Gütekriterien sowohl der BSS als auch der GAF nach erfolgter eingehender *inhaltlicher* Prüfung nicht für eine differenzierte Abbildung des Schweregrades in Frage kommen, da die Operationalisierung der drei verschiedenen Dimensionen des BSS und auch die Ankerpunkte der GAF-Skala sehr unscharf sind (vgl. Abbildung 11). Auch der kombinierte Einsatz, den Lange und Heuft (2002) vorschlagen, kann unter dem Gesichtspunkt der Praktikabilität nicht weiter verfolgt werden.

Da in Deutschland bislang noch kein Instrument existiert, das sowohl den Schweregrad einer psychischen Störung differenziert abbildet und gleichzeitig den Anforderungen an die Praktikabilität gerecht wird, wurden geeignete Instrumente zur Erfassung der Schwere einer psychischen Störung aus dem internationalen Bereich überprüft. Bei der Überprüfung, die ausführlich im Kapitel 3 dieser Arbeit vorgenommen wurde, erwies sich insbesondere der HoNOS als ein zentraler Prädiktor des Ressourcenverbrauchs.

# 6 Die Health of the Nation Outcome Scales (HoNOS): Ein Instrument zur differenzierten Erfassung der Schwere einer psychischen Störung

In Großbritannien wurde mit der Zielsetzung der differenzierten Erfassung des Schweregrades und der sozialen Funktionsfähigkeit ein Instrument („**Health of the Nation Outcome Scales, HoNOS“**) entwickelt, welches aber auch gleichzeitig im Routineeinsatz verwendbar sein sollte (Wing et al., 1998). Inzwischen wird dieses Instrument auch in anderen Ländern eingesetzt. In einem australischen Projekt zur Entwicklung eines Klassifikationssystems für Patienten mit psychischen Störungen erwies sich der HoNOS z.B. als bedeutsamer Prädiktor des Ressourcenverbrauches (Buckingham et al., 1998). Die Entwicklung des Instrumentes und die Ergebnisse zur Überprüfung der psychometrischen Eigenschaften werden im folgenden Abschnitt beschrieben.

## 6.1 Die Entwicklung der HoNOS in Großbritannien

Innerhalb der „Health of the Nation Strategy“ formulierte die Regierung Großbritanniens drei umfassende Ziele zur Verbesserung der Versorgung psychisch Kranker. Das erste Ziel bezog sich darauf, die Gesundheit und das soziale Funktionsniveau psychisch Kranker signifikant zu verbessern. Das Royal College of Psychiatrists´ wurde daraufhin 1992 vom Department of Health damit beauftragt ein Set von Skalen bereitzustellen, welches die Gesundheit und soziale Funktionsfähigkeit von Patienten mit psychischen Störungen erfassen sollte. Dabei sollte das Instrument verschiedene Anforderungen erfüllen. In erster Linie sollte es für den routinemäßigen Einsatz geeignet sein, sollte generelle klinische Probleme und soziale Funktionsfähigkeit messen, änderungssensitiv sein und über Reliabilitäts- und Validitätskennwerte verfügen (Wing et al., 1996; Wing et al., 1998). Entwickelt wurde das Instrument auch in Kooperation mit anderen Arbeitsgruppen (z.B. mit dem National Casemix Office, NCO oder mit der Schizophrenia Project of the Clinical Standards Advisory Group, CSAG) und mit Experten aus verschiedenen Berufsgruppen (u.a. Krankenpflegekräfte, Psychologen und Sozialarbeiter).

Zu Beginn der Entwicklung stand eine ausführliche Literaturrecherche zu Messinstrumenten psychologischer und sozialer Funktionsfähigkeit. Eine ausführliche Auflistung der verwendeten Literatur findet sich im Anhang des Reports von Wing et al. (1996). Es konnte aber kein Instrument identifiziert werden, dass kurz genug für einen routinemäßigen klinischen Einsatz war, landesweit verwendbar und gleichzeitig klinische und soziale Funktionsfähigkeit erfasste. Mit dem Ziel eine hohe Praktikabilität und routinemäßige Anwendbarkeit herzustellen, war

auch die Anforderung verbunden, möglichst Skalen zu konstruieren, die einfach und schnell zu erfassen waren. Vier Funktionsbereiche wurden dabei als zentral definiert: 1) Distress durch Depression, Ängstlichkeit, Halluzinationen oder Wahn, 2) Defizite in Basisfunktionen wie z.B. psychomotorische Verlangsamung, kognitive und physische Beeinträchtigungen, 3) Verhaltensprobleme wie z.B. Selbstverletzung oder Gewaltanwendung, 4) Probleme die mit der Umgebung des Patienten in Zusammenhang stehen wie z.B. finanzielle oder interpersonelle Probleme. Dabei nahm die differenzierte Erfassung der Schwere einer psychischen Erkrankung bei der Skalenkonstruktion einen hohen Stellenwert ein.

Die Entwicklung der HoNOS durchlief 4 sorgfältig geplante und durchgeführte Phasen in einem Zeitraum zwischen 1992 bis 1995 (Wing et al., 1998). Zunächst entstand als Ergebnis einer Literaturanalyse und Expertenbefragungen ein 20 Item umfassendes Instrument (HoNOS-1). Die 20 Items wurden von den Testautoren aufgrund inhaltlicher Kriterien 4 Skalen zugeordnet („Verhalten“, „Impairment“, „Symptome“ und „Soziale Funktionsfähigkeit“). Daneben besteht auch die Möglichkeit, die Items zu einem Gesamtscore aufzuaddieren. Die HoNOS-1 wurde dann in einer ersten Pilotphase hinsichtlich ihrer Akzeptabilität, Struktur und Sensitivität anhand einer Stichprobengröße von n = 152 Patienten aus 9 verschiedenen Einrichtungen getestet und anschließend die Itemanzahl von 20 auf 12 reduziert (HoNOS-2). HoNOS-2 wurde dann wiederum an einer Stichprobengröße von n = 100 Patienten aus 7 verschiedenen Einrichtungen hinsichtlich Akzeptabilität, Struktur und Sensitivität überprüft. Die Ergebnisse des 2. Pilottests unterschieden sich nicht wesentlich von den Ergebnissen der 1. Pilotphase. Aufgrund der bis dahin gewonnenen Erfahrungen wurde HoNOS-3 konstruiert und vor allem dem Instrument ein umfangreiches, detailliertes Glossar hinzugefügt. In der sich anschließenden Feldstudie mit insgesamt 25 beteiligten Einrichtungen (N = 2.706 Patienten) wurde der HoNOS-3 bezüglich seiner Akzeptabilität und seiner psychometrischen Gütekriterien untersucht. Es wurde u.a. auch eine Überprüfung der faktoriellen Validität der HoNOS-3 mittels explorativer Faktorenanalyse mit Varimax-Rotation vorgenommen. Die Analyse ergab eine 5-Faktorenlösung mit 63% Varianzaufklärung, was die Testautoren als relativ nah angelehnt an der postulierten 4-Faktorenlösung werteten. Aufgrund zahlreicher Anregungen von klinischen Experten unter Nutzung unterschiedlicher Quellen (z.B. Expertenbefragungen, Reaktionen auf den HoNOS-Newsletter) nahmen Wing et al. (1998) geringfügige Ergänzungen und Änderungen der HoNOS-3 Version vor und arbeiteten diese in die Endversion ein (HoNOS-4).

## 6.2 Zentrale psychometrische Kennwerte der HoNOS in Großbritannien

Die Endversion wurde in 6 Einrichtungen anhand einer Gesamtstichprobe von N = 754 Patienten eingesetzt und bezüglich ihrer psychometrischen Eigenschaften analysiert. Hauptsächlich waren Patienten mit einer psychotischen Störung als Erstdiagnose (44,3%, ICD-10), gefolgt von Patienten mit affektiven Störungen (33,3%) in der Stichprobe vertreten. Generell lässt sich sagen, dass ein hoher Anteil von Patienten über 65 Jahre in die Stichprobe eingeschlossen war (31,7% über 65 Jahre).

Zur Überprüfung der Reliabilität der HoNOS-Endversion wurden Interraterkorrelationen bestimmt. In der Hauptphase der Erhebung wurden n = 197 Patienten durch die Trainer vor Ort und durch eine therapeutische Bezugsperson in den Skalen der HoNOS eingeschätzt. Die Intraklassenkoeffizienten (ICC) waren für die Mehrzahl der 12 Items zufrieden stellend (ICC = .61 - .99), für Item 8 („Andere Symptome“) und Item 12 („Probleme durch die Bedingungen im Beruf und im Alltag“) konnten allerdings nur moderate Werte erreicht werden (ICC = .52, ICC = .49). Diese Ergebnisse könnten eventuell mit der Stichprobenzusammensetzung in Verbindung stehen. Z.B. wäre bei Item 12 möglich, dass gerade die älteren Patienten in der Stichprobe, welchen keinen Beruf mehr ausübten, auch in ihren allgemeinen beruflichen und alltäglichen Bedingungen schwerer einschätzbar waren, was zur Inkonsistenz zwischen den Ratern geführt haben könnte. Tabelle 8 bietet einen Überblick über die Werte der Intraklassenkoeffizienten für die HoNOS Items in zwei Studien (Wing et al., 1998).

Zur Überprüfung der Validität des Instrumentes wurden zwei Fremdeinschätzungsinstrumente, die Brief Psychiatric Rating Scale (BPRS) und die Role Functioning Scale (RFS)[9], hinzugezogen. Dabei zeigte sich ein deutlicher Zusammenhang des HoNOS Gesamtscores sowohl zum Gesamtwert des BPRS (r = .84) als auch zu den vier Skalenwerten der RFS (r = .52, .64, .67 und .73) (Wing et al., 1998).

---

[9] Interraterreliabilitäten für den Gesamtwert der BPRS liegen zwischen .87 und .97 (Collegium Internationale Psychiatriae Scalarum (CIPS), 1996a) und können nach den Kriterien von Landis und Koch (1977) als zufrieden stellend bewertet werden. Interraterreliabilitäten für die vier Skalen der RFS liegen zwischen .64 und .82 und können als zufrieden stellend bewertet werden (Goodman et al., 1993).

Tabelle 8: Interraterkorrelationen der HoNOS-Items in 2 Studien (Nottingham und Manchester) (Wing et al., 1998)

| Items der HoNOS | Interraterkorrelationen der Studie in Nottingham (n = 100) | Interraterkorrelationen der Studie in Manchester (n = 97) |
|---|---|---|
| 1. Aggressivität | .97 | .80 |
| 2. Selbstverletzungen | .88 | .92 |
| 3. Substanzmissbrauch | .99 | .61 |
| 4. Kognitive Probleme | .81 | .92 |
| 5. Körperliche Erkrankung | .88 | .89 |
| 6. Halluzinationen und Wahn | .87 | .92 |
| 7. Gedrückte Stimmung | .84 | .89 |
| 8. Andere Symptome | .95 | .52 |
| 9. Beziehungsprobleme | .74 | .78 |
| 10. Alltägliche Aktivitäten | .71 | .90 |
| 11. Wohnbedingungen | .83 | .72 |
| 12. Beruf/Freizeit | .49 | .51 |

**Anmerkungen:** *HoNOS = Health of the Nation Outcome Scales; n = Stichprobengröße*

### 6.3 Zentrale empirische Befunde zu psychometrischen Eigenschaften der HoNOS

Nachdem nun die psychometrischen Eigenschaften der Endversion aus der Studie der Testautoren (Wing et al., 1998) vorgestellt wurden, sollen im Folgenden empirische Befunde aus verschiedenen Studien[10] einen Beitrag zur **Diskussion der Kennwerte** der HoNOS leisten. Die meisten Untersuchungen stammen aus Großbritannien und Australien, da in diesen beiden Ländern das Instrument innerhalb der nationalen Gesundheitsstrategie als „Minimum Data Set“ eingesetzt wird. Bevor die einzelnen Gütekriterien (Praktikabilität, Reliabilität und Validität) näher beleuchtet werden, soll zunächst eine Beschreibung des Settings und der Stichproben der ausgewählten Studien erfolgen.

[10] Eingeschlossen wurden hierbei zehn Studien, die sich als unmittelbar relevant für die zu beantragende Studie erwiesen haben (Goldney et al., 1998; Orrell et al., 1999; Trauer et al., 1999; Adams et al., 2000; Brooks, 2000; Browne et al., 2000; McClelland et al., 2000; Audin et al., 2001; Page et al., 2001; Parker et al., 2002).

Die Studien wurden in **unterschiedlichen Versorgungsbereichen** für Patienten mit psychischen Störungen durchgeführt. Untersucht wurden Patienten, die sich in stationärer psychiatrischer/psychotherapeutischer Behandlung (Goldney et al., 1998; Orrell et al., 1999; Trauer et al., 1999; Trauer, 1999; Adams et al., 2000; Brooks, 2000; McClelland et al., 2000; Page et al., 2001), in teilstationärer Behandlung, z.B. Tageskliniken (Orrell et al., 1999; McClelland et al., 2000) und in ambulanter psychotherapeutischer Behandlung (Orrell et al., 1999; Browne et al., 2000; McClelland et al., 2000; Audin et al., 2001; Parker et al., 2002) befanden. Am häufigsten findet das Instrument zur differenzierten Erfassung des Schweregrades bei Patienten mit „Substanz- oder Alkoholabusus“ und „Schizophrene Störung“ (F1, F2 Kapitel der ICD-10) Verwendung (Orrell et al., 1999; Trauer et al., 1999; Adams et al., 2000; Brooks, 2000; Browne et al., 2000; McClelland et al., 2000; Parker et al., 2002), es liegen jedoch auch einige wenige empirische Ergebnisse für andere Diagnosegruppen des F Kapitels der ICD-10 vor (z.B. für Affektive Störungen Goldney et al., 1998; Audin et al., 2001; Page et al., 2001). Hinsichtlich der Studiendesigns wird eine große Spannbreite der Stichprobenumfänge deutlich (N = 100 - 2.137). Zur Stichprobenzusammensetzung lässt sich sagen, dass in den meisten hier aufgeführten Studien ein leicht höherer Anteil an Männern zu verzeichnen war, und dass das durchschnittliche Alter der Patienten zwischen 35 bis 43 Jahre varrierte (SD = 9 - 17 Jahre).

Hauptanliegen von Wing et al. (1996) war es, ein Instrument zu konstruieren, das sich im routinemäßigen Einsatz bewährt und den Anforderungen der **Praktikabilität** genügt. Erste empirische Hinweise dazu lieferte die Studie von Browne et al. (2000). Die Therapeuten der Studie benötigten während der routinemäßigen Anwendung durchschnittlich 2 bis 3 min zum Ausfüllen, womit die Ergebnisse der Testautoren (Wing et al., 1998) hinsichtlich zufrieden stellender Praktikabilitätswerte der HoNOS repliziert werden konnten.

Um Hinweise auf die Güte der **Reliabilität** eines Fremdeinschätzungsinstrumentes zu gewinnen, ist die Prüfung von Interraterübereinstimmungen notwendig (Orrell et al., 1999; Trauer et al., 1999; Brooks, 2000). Eine zweite Überprüfung richtet sich auf die Homogenität der Skalen der HoNOS und in wie weit sich die postulierte weitest gehende Unabhängigkeit der Skalen replizieren lässt (McClelland et al., 2000).

Trauer et al. (1999) und Orrell et al. (1999) konnten für eine Großteil der Items zufrieden stellende **Interraterkoeffizienten** aufzeigen. In diesen beiden Studien ergaben sich allerdings

auch erste Hinweise darauf, dass die Items 11 und 12, welche allgemeine Lebensbedingungen des Patienten erfassen, eher niedrigere Interraterübereinstimmungen hatten. In der Studie von Brooks (2000) wiesen die meisten der 12 Items (7 von 12), verglichen mit den Ergebnissen von Wing et al. (1996) und Trauer et al. (1999), keine zufrieden stellenden Koeffizienten auf.

Ein Reliabilitätsindikator der HoNOS ist die Überprüfung der Homogenität der Skalen. **Itemanalytische Auswertungen** von McClelland et al. (2000) und Audin et al. (2001) zeigten heterogene Ergebnisse. Audin et al. (2001) fanden nur für wenige Items der HoNOS akzeptable Verteilungseigenschaften, was vermutlich stichprobenabhängig war. McClelland et al. (2000) konnten aber die von den Testautoren postulierte, weitest gehende Unabhängigkeit der 12 Items untereinander replizieren (2 Interkorrelationen über r = .40). Im Artikel von McClelland et al. (2000) fand sich allerdings keine Angabe, wie viele Interkorrelationen über .30 lagen.

Zur Sicherstellung und Erhöhung der Reliabilität entwickelten Wing et al. (1996) ein **Schulungskonzept** im HoNOS. Empirische Befunde belegen allerdings, dass die Höhe der Interraterreliablitäten nicht notwendigerweise von einer Schulung abhängt. So untersuchten Rock und Preston (2001) HoNOS trainierte und nicht-trainierte Gruppen und kamen zu dem Schluss, dass weder klinische Erfahrung noch eine Schulung einen signifikanten Unterschied im Ausmaß der Interraterübereinstimmung bewirkte. Das lässt darauf schliessen, dass auf ein Training der Beurteiler eher verzichtet werden könnte, wobei dazu noch weitere Studien ausstehen.

Zur Überprüfung der **faktoriellen Validität** des Instruments wurden von einigen Autoren Faktorenanalysen und interne Konsistenzen der Skalen der HoNOS berechnet. So berichteten Page et al. (2001) eine interne Konsistenz des Gesamtwertes von $\alpha$ = .59, was sicherlich vor dem Hintergrund der Annahme der weitestgehenden Unabhängigkeit der 12 Items zu erwarten ist. Mithilfe einer Hauptkomponentenanalyse versuchten McClelland et al. (2000) die vier Skalen der HoNOS („Verhalten“, „Symptome“, „Impairment“ und „soziale Funktionsfähigkeit“) zu replizieren. Die Autoren konnten 4 unabhängige Faktoren mit einer Gesamtvarianzaufklärung von 55% extrahieren. Der erste Faktor „Schwere der Erkrankung“ beinhaltete dabei 8 der 12 HoNOS-Items. Item 2 „Selbstverletzung“ und Item 7 „Gedrückte Stimmung“ wiesen Ladungen auf dem zweiten Faktor auf, Item 5 „Probleme mit körperlichen Erkrankungen“ bildete den 3. Faktor und Item 8 „Andere Symptome“ wurde ebenfalls als alleiniger Fak-

tor identifiziert. McClelland et al. (2000) konnten damit nicht die 4-Faktorenstruktur von Wing et al. (1998) replizieren. Auch Trauer (1999) untersuchte die faktorielle Validität der HoNOS anhand einer Stichprobe von 2.137 Patienten aus fünf psychiatrischen Krankenhäusern. So konnte er für den Gesamtscore eine zufrieden stellende interne Konsistenz mit $\alpha = .75$ aufzeigen, allerdings berichtete er für die 4 Subskalen der HoNOS eine Spannweite der internen Konsisten von $\alpha = .28$ - $\alpha = .71$. Im nächsten Schritt überprüfte Trauer (1999) mittels einer konfirmatorischen Faktorenanalyse die von Wing et al. (1998) postulierte 4-Faktorenstruktur der HoNOS. Im Einklang mit den Befunden von McClelland et al. (2000) konnte auch Trauer (1999) die 4-Faktorenstruktur der Testautoren nicht zufrieden stellend replizieren. Diese empirischen Befunde zur moderaten internen Konsistenz der 4 Subskalen sind vor dem Hintergrund der weitestgehenden Unabhängigkeit der 12 Einzelitems zu erwarten. Dieser Aspekt hat aber auch Auswirkungen auf die Verwendung von aggregierten Daten wie z.B. des Gesamtscores. So schlussfolgert Trauer (1999): „The level of aggregation of individual items scores will depend upon the uses that the scale will be put. In general, it is likely that those higher in the organisational structure will prefer aggregation (total scores) while those lower, and who are engaged with direct client contact, will prefer disaggregation (individual item scores).” (zitiert nach Trauer, 1999, S. 506).

Zur Überprüfung der **Validität** der HoNOS wurde untersucht, in wie weit sich hinsichtlich des Schweregrades erwartungsgemäß Unterschiede zwischen dem Alter, dem Geschlecht und der Diagnose von Patienten nachweisen lassen. So fanden Trauer et al. (1999), dass ein jüngeres Alter der Patienten signifikant mit den Items Aggressivität, Selbstverletzung und Alkohol- und Drogenkonsum assoziiert war, ein höheres Alter dagegen eher mit den Items kognitive Probleme, Probleme durch körperliche Erkrankung und Behinderung einherging. Erwartungsgemäß zeigte sich weiterhin, dass Frauen höher auf den Items Selbstverletzung, Probleme durch körperliche Erkrankung und depressiver Stimmung scorten, dagegen Männer bei den Items Alkohol- und Drogenkonsum, Halluzinationen und Probleme durch Wohn- und Arbeitsbedingungen höhere Werte einnahmen (Trauer et al., 1999). Des Weiteren zeigte sich, dass Patienten mit einer depressiven Erkrankung höhere Werte im Depressionsitem der HoNOS, im Item „andere Symptome“, beim Item Probleme mit Beziehungen und Probleme durch körperliche Erkrankung aufwiesen. McClelland et al. (2000) fanden in der Gruppe der Patienten mit Substanzabusus die höchsten Werte im Gesamtscore der HoNOS-4 (Gesamtscore = Summenwert über alle 12 Items), hohe Werte nahmen auch Patienten mit einer schizophrenen oder bipolar-affektiven Störung ein, gefolgt von Patienten mit depressiver Erkran-

kung, Persönlichkeitsstörungen, neurotischen und Ess-, Schlaf- und Stressstörungen. Dieses Ergebnis zeigte sich auch in der Untersuchung von Browne et al. (2000).

Zur weiteren Überprüfung der Validität der HoNOS wurden verschiedene fremd- und selbsteingeschätzte Instrumente eingesetzt: die Brief Psychiatric Rating Scale (Overall & Gorham, 1976), die Global Assessment of Functioning Scale (Endicott et al., 1976), die Short-Form 36 (Ware, 2000) und die Symptom-Check-List 90R (Derogatis, 1977). Auch hier sollen im Folgenden nur die für die vorliegende Untersuchung relevanten Ergebnisse vorgestellt werden. Mehrere Studien untersuchten den Zusammenhang zwischen dem HoNOS Gesamtscore und dem BPRS Gesamtscore. So konnten McClelland et al. (2000) und Orrell et al. (1999) einen signifikant mittleren Zusammenhang zu Beginn der Behandlung und einen signifikant höheren Zusammenhang zwischen den Gesamtscores beider Instrumente nach der Behandlung aufzeigen. Auch Adams et al. (2000) fanden zwar eher niedrigere, aber auch signifikante Korrelationen ($r \sim .30$). Ebenso signifikant mittlere bis höhere Zusammenhänge zu Beginn und am Ende der Behandlung ergaben sich zwischen dem HoNOS Gesamtscore und der GAF-Skala (Orrell et al., 1999; Browne et al., 2000; McClelland et al., 2000; Parker et al., 2002).

Bei den Zusammenhängen der HoNOS-4 mit selbst eingeschätzten Instrumenten zeigten sich eher widersprüchliche Ergebnisse. Während Adams et al. (2000) signifikante, wenn auch eher niedrigere Korrelationen ($r = .25$) zwischen HoNOS Gesamtscore und dem Global Severity Index der SCL-90R feststellte, berichtete Brooks (2000) keine signifikanten Zusammenhänge zu Beginn und am Ende der Behandlung. Genauso verhält es sich auch mit der SF-36. Brooks (Brooks, 2000) fand auch hier keine signifikanten Korrelationen zwischen dem HoNOS Gesamtwert und der psychischen Summenskala des SF-36 zu Beginn und am Ende einer psychiatrisch/psychotherapeutischen Behandlung, konnte allerdings signifikante Korrelationen zwischen der körperlichen Summenskala des SF-36 und dem HoNOS Gesamtwert aufzeigen. Dagegen zeigten Orrell et al. (1999) signifikante Zusammenhänge zwischen dem Gesamtscore im HoNOS und der SF-36 auf.

Des Weiteren konnten Trauer et al. (1999) zwar eine Abnahme der Werte im HoNOS Gesamtscore zwischen Beginn und Ende einer stationären psychiatrisch/psychotherapeutischen Behandlung aufzeigen, jedoch war dieses Ergebnis nicht signifikant. Trauer et al. (1999) schlussfolgerten daraus, dass eventuell das Intervall zwischen den beiden Messzeitpunkten zu kurz gewesen sein könnte (im Mittel 28 Tage). Dagegen berichteten Audin et al. (2001) eine

signifikante Abnahme des HoNOS Gesamtscores mit einem mittleren Effekt von d = .69. Nur drei Items der HoNOS wiesen keine signifikanten Veränderungen zwischen Beginn und Ende der ambulanten psychotherapeutischen Behandlung auf. Zu den gleichen Ergebnissen kamen Goldney et al. (1998) und Brooks (2000). In der Studie von Goldney et al. (1998) wurde für den HoNOS Gesamtscore ein großer Effekt mit d = .92 erreicht. Vergleiche zwischen der HoNOS und der SF-36 in der Studie von Page et al. (2001) ergaben, dass der HoNOS insgesamt eher Veränderungen nach der Behandlung anzeigte bei einer durchschnittlichen Behandlungsdauer von 10 Tagen als die SF-36 in einer privaten psychiatrischen Klinik. So variierten die Effektstärken im HoNOS in einem Bereich zwischen 0.28 - 1.56, wogegen die Effektstärken im SF-36 Werte zwischen 0.01 - 0.53 einnahmen. Einschränkend muss noch hinzugefügt werden, dass bei den gerade berichteten Ergebnissen die Einschätzung im HoNOS nicht unabhängig von der Person des Therapeuten erfolgte, was sich sicherlich auf die Güte der Beurteilung am Ende der Therapie auch im Sinne der Überschätzung des Therapieerfolgs auswirken kann.

**Zusammenfassende Bewertung**

**Zusammenfassend** lässt sich festhalten, dass die empirischen Befunde zu den psychometrischen Eigenschaften der HoNOS heterogen sind. Es gibt erste Hinweise zur Praktikabilität des Instrumentes bei bestimmten Patientengruppen, wie z.B. Patienten mit schizophrenen Störungen oder bei Patienten mit Substanzabusus. Forschungsbedarf besteht allerdings noch bei Patientengruppen, die zu den F3 - F6 Kapiteln der ICD-10 zu rechnen sind, z.B. Patienten mit einer Angst- oder Zwangsstörung als Hauptdiagnose. Auch die Ergebnisse zur Reliabilität des Instrumentes variieren sehr stark, womit auch hier ein weiterer Forschungsbedarf gekennzeichnet ist. Betrachtet man die Ergebnisse zur Validität so zeichnet sich auch hier ein eher heterogenes Bild ab, wobei es zurzeit noch vor allem an Untersuchungen zu Zusammenhängen mit selbst eingeschätzten Instrumenten mangelt.

**Insgesamt betrachtet** fehlt es derzeit an weiteren Studien zu psychometrischen Eigenschaften der HoNOS an Patienten vor allem auch aus **anderen Versorgungsbereichen** außerhalb der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie. Die vorliegende Untersuchung ist in einem Versorgungsbereich der stationären Rehabilitationsbehandlung von

Patienten mit psychischen/psychosomatischen Störungen angesiedelt. Wie Schulz & Koch (2002) in einer Expertise zur stationären psychosomatischen/psychotherapeutischen Versorgung in Norddeutschland aufzeigen konnten, besteht ein großer Überschneidungsbereich zwischen den unterschiedlichen Versorgungsbereichen von Patienten mit psychsichen Störungen (vgl. hierzu ausführlich Kapitel 1). Das Instrument wurde in Großbritannien vorrangig mit der Intention entwickelt, den Schweregrad von Patienten mit psychischen Störungen *in allen Versorgungsbereichen* differenziert und praktikabel zu erfassen. Inwieweit sich nun der Einsatz der HoNOS-D (deutsche Version der HoNOS) im nationalen Versorgungsbereich der stationären Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen innerhalb eines Klassifikationssystems als geeignet erweist, soll durch die vorliegende Untersuchung überprüft werden.

**Erwartet wird**, das die psychometrischen Analysen einen Beitrag zur Weiterentwicklung eines validen Screeninginstrumentes leisten, woran es derzeit für Patienten mit psychischen Störungen noch *in allen Versorgungsbereichen* fehlt. Der spätere Nutzen eines solchen validen Screeninginstruments wäre vielfältig. Es wird erwartet, dass der HoNOS-D einen Beitrag zur Fallgruppenbildung für Patienten mit psychischen Störungen in stationärer Behandlung leistet und somit auch übergreifend zum internen Ressourcenmanagement in den Kliniken verwendbar sein kann. Weiterhin könnte eine praktikable, differenzierte und valide Erfassung des Schweregrads von Patienten mit psychischen Störungen ein Instrument zur differentiellen Indikationsstellung für psychotherapeutische Behandlungen darstellen, auch im Rahmen eines Klassifikationsinstrumentes.

Im nächsten Kapitel sollen der Übersetzungsprozess der HoNOS in die deutsche Sprache und die deutsche Version der HoNOS (HoNOS-D) dargestellt werden.

# 7 Die deutsche Version der HoNOS (HoNOS-D)

## 7.1 Der Übersetzungsprozess der HoNOS in die deutsche Sprache

In Anlehnung an das von der Weltgesundheitsorganisation empfohlene Vorgehen (Sartorius & Kuyken, 1994) wurde im Februar 2003 in einem ersten Schritt das Score-Sheet und das umfangreiche Glossar der HoNOS in einem expertengestützten Konsensusverfahren von drei Mitgliedern der Arbeitsgruppe für Psychotherapie- und Versorgungsforschung (AGPV) am Institut und Poliklinik für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf von der englischen in die deutsche Sprache übersetzt. Dieser Schritt erfolgte bereits in Autorisierung durch die Testautoren der englischen Originalversion des Royal College of Psychiatrist´s in London (Großbritannien). Im nächsten Schritt wurde die deutsche Übersetzung von einem unabhängigen Experten - ebenfalls mit langjähriger klinischer Erfahrung - begutachtet. Seine Korrekturen wurden dann mit den Übersetzern der AGPV diskutiert und eingearbeitet. Anschließend wurde diese erste deutsche Version von einem „Native Speaker“ zurückübersetzt und den Testautoren der Originalfassung (vertreten durch Prof. Lelliot und Mr. James) zum Abgleich mit der englischen Originalversion der HoNOS (Endversion) zur Korrektur übersandt. Nach geringfügigen Veränderungen und Anmerkungen durch die Testautoren erfolgte die Fertigstellung der endgültigen Fassung des Score-Sheets und des Glossars der deutschen Version der HoNOS-D im April 2003. Der Übersetzungsprozess ist graphisch in Abbildung 12 veranschaulicht.

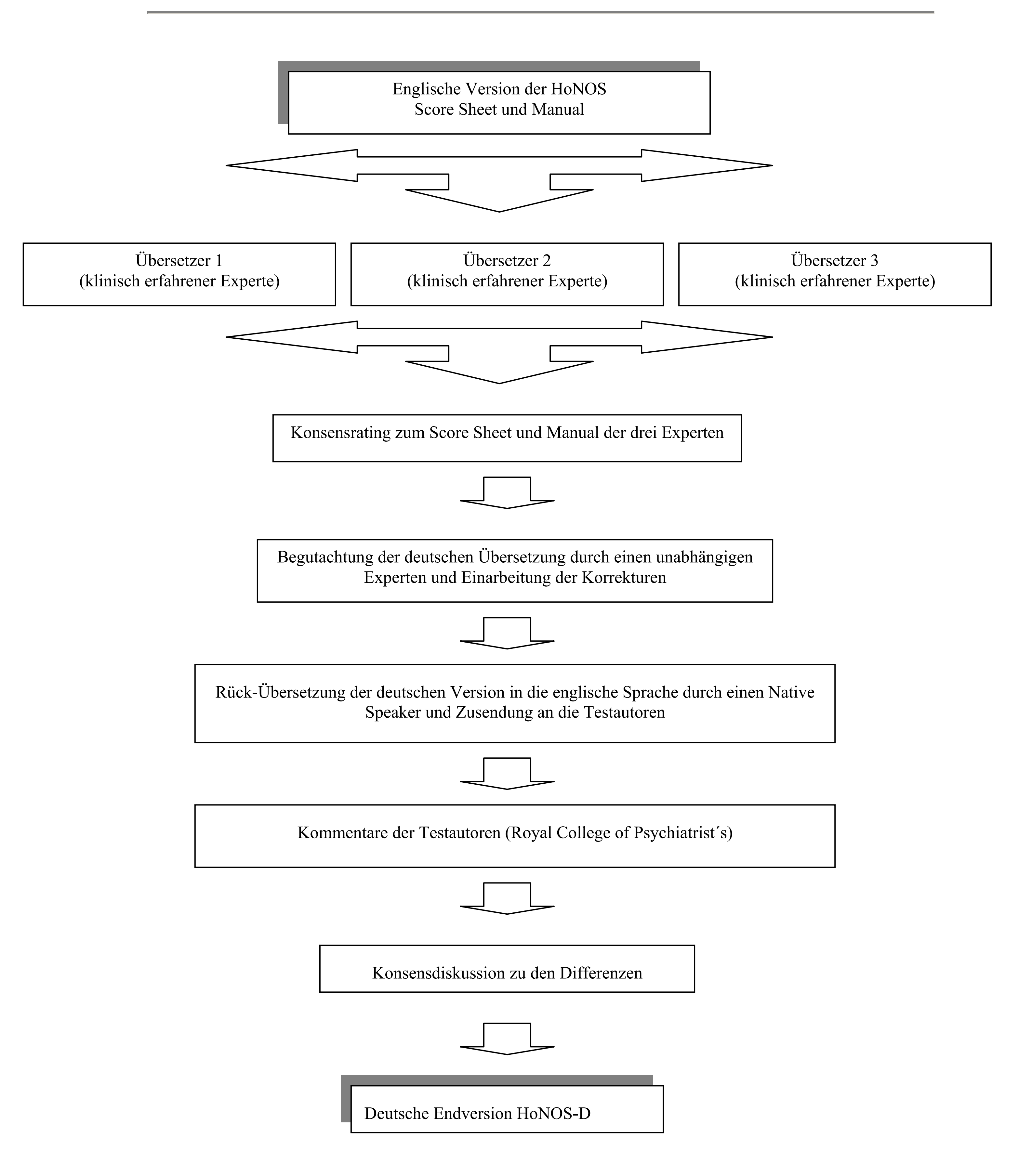



Abbildung 12: Übersetzungsprozess (autorisiert) der „Health of the Nation Outcome Scales“ in die deutsche Sprache (HoNOS-D)

### 7.2 Aufbau der HoNOS-D

Bei der HoNOS-D handelt sich um ein international weit verbreitetes Fremdeinschätzungsinstrument. Die empirischen Befunde zu psychometrischen Eigenschaften der HoNOS (Originalversion) wurden bereits im Einleitungsteil (vergleiche Kapitel 6, Abschnitt 6.2 und 6.3) der vorliegenden Arbeit dargelegt.

Die HoNOS-D ist ein 12 Item umfassendes Instrument zur Erfassung des differenzierten Schweregrades einer psychischen Erkrankung (vgl. Anhang 16.1.1 und Tabelle 9). Jedes der 12 Items beschreibt einen spezifischen Problembereich von Patienten mit psychischen Störungen, z.B. Item 7 „Gedrückte Stimmung“. Dem Beurteilungsformat liegt eine 5-stufige Likert-Skala von *0 = kein Problem* bis *4 = sehr schweres Problem* zugrunde. Die 12 Items können auch zu vier Skalen „Verhalten“, „Impairment“, „Symptome“ und „soziale Funktionsfähigkeit“ zusammengefasst werden (siehe Tabelle 9). Darüber hinaus ist es möglich einen Gesamtscore über alle 12 Items zu bestimmen. Hierzu sollten mindestens 10 der 12 Items beantwortet sein.

Tabelle 9: Die vier Dimensionen der HoNOS-D mit den dazugehörigen Items

| Dimensionen der HoNOS-D | Items | Itembezeichnung der Dimensionen (Kurzbezeichnung) |
|---|---|---|
| Verhalten | Item 1 | Überaktives, aggressives, Unruhe stiftendes Verhalten (Aggression) |
| | Item 2 | Absichtliche Selbstverletzung (Selbstverletzung) |
| | Item 3 | Problematischer Alkohol- oder Drogenkonsum (Substanzmissbrauch) |
| Impairment | Item 4 | Kognitive Probleme |
| | Item 5 | Probleme in Zusammenhang mit körperlicher Erkrankung oder Behinderung (Probleme mit körperlichen Erkrankungen) |
| Symptome | Item 6 | Probleme in Zusammenhang mit Halluzinationen oder Wahnvorstellungen (Halluzinationen und Wahnvorstellungen) |
| | Item 7 | Gedrückte Stimmung |
| | Item 8 | Andere psychische und verhaltensbezogene Probleme (Andere Symptome) |
| Soziale Funktionsfähigkeit | Item 9 | Probleme mit Beziehungen |
| | Item 10 | Probleme mit alltäglichen Aktivitäten |
| | Item 11 | Probleme durch die Wohnbedingungen |
| | Item 12 | Probleme durch die Bedingungen im Beruf und im Alltag |

Zur Konkretisierung der Schweregradeinschätzungen gibt ein umfangreiches Glossar detaillierte Anweisungen (vgl. Anhang 16.1.2). Um die differenzierten Operationalisierungen zu demonstrieren, ist in Abbildung 13 ein Beispielitem (Item 7 „Gedrückte Stimmung") aus dem Glossar aufgeführt. So wird z.B. bei Item „Gedrückte Stimmung“ der Rater angewiesen, eine „1“ (klinisch unbedeutendes Problem ohne Handlungsbedarf) zu vergeben, wenn bedrückte oder geringe Veränderungen in der Stimmung des Patienten bestehen. Dagegen soll der Rater eine „4“ (schweres bis sehr schweres Problem) angeben, wenn eine schwere oder sehr schwere Depression mit Schuld oder Selbstanklage vorliegt. Des Weiteren wird für jedes Item spezifisch angegeben, welche Verhaltensweisen bei diesem Item nicht mit in die Beurteilung einzubeziehen sind (vgl. Abbildung 13 und Anhang 16.1.2).

**Item 7 „Gedrückte Stimmung“**

** Beziehen Sie Überaktivität oder Agitiertheit, welche auf der Skala 1 eingeschätzt werden,* ***nicht*** *mit ein.*

** Beziehen Sie Suizidgedanken oder –versuche, welche auf der Skala 2 eingeschätzt werden,* ***nicht*** *mit ein.*

** Beziehen Sie Wahnvorstellungen oder Halluzinationen, welche auf der Skala 6 eingeschätzt werden,* ***nicht*** *mit ein.*

0 Kein Problem in Zusammenhang mit gedrückter Stimmung während des Bewertungszeitraums.

1 Bedrückt; oder geringe Veränderungen in der Stimmung.

2 Leichte(s), jedoch eindeutig vorhandene(s) Depression und Leiden (z. B. Schuldgefühle; Verlust des Selbstwertgefühls).

3 Depression mit unangemessenen Selbstvorwürfen; zwanghaft beschäftigt mit Schuldgefühlen.

4 Schwere oder sehr schwere Depression mit Schuld oder Selbstanklage.

Abbildung 13: Beispielitem (Item 7) der HoNOS-D („Health of the Nation Outcome Scales“, deutsche Version) zur Illustration der Operationalisierungen der einzelnen Skalenstufen

Nach der Darstellung des Übersetzungsprozesses der HoNOS in die deutsche Sprache und der Charakterisierung des Instrumentes, sollen nun im nächsten Kapitel die vor dem Hintergrund der Literaturübersicht und empirischen Befunde zu internationalen und nationalen Klassifikationssystemen abgeleiteten Fragestellungen und Hypothesen formuliert werden.

# 8 Fragestellungen und Hypothesen

## 8.1 Fragestellungen

Vor dem Hintergrund der in Kapitel 3 und 4 dargestellten Befunde aus internationalen zu allgemeinen DRG-Systemen komplementär, entwickelten Klassifikationssystemen (Buckingham et al., 1998; Gaines et al., 2003; Horn et al., 1989; Hirdes et al., 2003; Ashcraft et al., 1989) und der nationalen Studien zum Ressourcenverbrauch in stationärer Behandlung (Scheytt et al., 1996; Zielke et al., 1997; Jerschke et al., 1998; Borgart & Meermann, 1999; Kächele für die Studiengruppe MZ-ESS, 1999; Richter, 1999, 2001; Kluge et al., 2002; Krautgartner et al., 2002; Mestel et al., 2002; Spießl et al., 2002; Spitzer et al., 2002; Roick et al., 2002a; Roick et al., 2002b) kann geschlussfolgert werden, dass der differenzierten Erfassung der Schwere einer psychischen Erkrankung innerhalb eines Klassifikationssystems eine zentrale Stellung zukommt. Wie in Kapitel 5 weiter dargelegt wurde, ist hierbei unter Umständen die Fremdeinschätzung des Schweregrades von besonderer Bedeutung. Nach Überprüfung bereits im deutschsprachigen Raum gut etablierter Fremdeinschätzungsinstrumente war festzustellen, dass bislang noch kein geeignetes deutschsprachiges Instrument zur differenzierten Erfassung des Schweregrades für alle Versorgungsbereiche existiert. In den komplementär zu den allgemeinen DRG-Systemen entwickelten internationalen Klassifikationssystemen für den Bereich psychischer Störungen (MH-CASC, NZ-CAOS und MHG, vgl. Kapitel 3) erwies sich jedoch ein Fremdeinschätzungsinstrument, die „Health of the Nation Outcome Scales HoNOS“, als besonders ressourcenverbrauchsrelevant. Hierzu wurde in Kapitel 6 allerdings dargelegt, dass die empirischen Befunde zu den psychometrischen Eigenschaften der HoNOS insgesamt als heterogen zu bewerten sind.

Ziel des empirischen Teils der vorliegenden Untersuchung ist es deswegen, zum einen die generelle Güte der deutschen Version der HoNOS (HoNOS-D) hinsichtlich ihrer psychometrischer Eigenschaften zu untersuchen und zum anderen die spezifische Eignung des Instrumentes bezüglich der Bildung von Fallgruppen für Patienten mit psychischen Störungen zu überprüfen. Dazu sollen folgende zentrale Fragestellungen beantwortet werden:

1. **In wie weit erweist sich die HoNOS-D generell bezüglich ihrer psychometrischen Eigenschaften als ein geeignetes Instrument?**

2. **In wie weit erweist sich die HoNOS-D als Instrument zur Fallgruppenbildung innerhalb eines Klassifikationssystems für Patienten mit psychischen Störungen als geeignet?**

Ein weiterer Befund der Literaturanalyse der vorliegenden Arbeit ist, dass die internationalen komplementär zu allgemeinen DRG-Systemen entwickelten Klassifikationssysteme zur Fallgruppenbildung vorrangig fremd eingeschätzte Merkmale verwenden. In der vorliegenden Arbeit wurde allerdings geschlussfolgert, dass sowohl selbst- wie auch fremd eingeschätzte Instrumente verwendet werden sollten. Deswegen soll als weitere Nebenfragestellung mit eher explorativen Charakter folgende Fragestellung bearbeitet werden:

3. ***In wie weit unterscheiden sich Fremd- und Selbsteinschätzungsverfahren hinsichtlich ihrer Bedeutsamkeit bei der Fallgruppenbildung von Patienten mit psychischen Störungen?***

## 8.2 Ableitung der Hypothesen

Bezogen auf die Hauptfragestellung der *generellen Eignung der HoNOS-D* sollen nun im Folgenden Hypothesen formuliert werden. Dabei soll die Testung auf Einzelitemebene durchgeführt werden, da die empirischen Ergebnisse zur Faktorenstruktur der HoNOS nur moderate Kennwerte ergaben (vgl. hierzu auch Kap. 6, Abschnitt 6.3). Weiterhin soll aber auch der Gesamtscore, gebildet über alle 12 Items, aufgrund inhaltlicher Überlegungen mit in die Testung einbezogen werden.

Auf Einzelitemebene (Hypothese 1) werden keine isolierten Hypothesen formuliert, sondern es soll hinsichtlich einer orientierenden Testung geprüft werden, in wie weit sich die in den Hypothesen postulierten Zusammenhänge aufzeigen lassen. Generell findet kein strenges hypothesenprüfendes Vorgehen im engeren Sinne statt, also in dem Sinne, dass Hypothesen aus einer Theorie abgeleitet und getestet werden, um diese Theorie zu falsifizieren oder zu

verifzieren. Bestimmte Erwartungen liegen aber aufgrund theoretischer Überlegungen und empirischer Befunde nahe, die bei den jeweiligen Hypothesen begründet werden. Wenn im Folgenden von Hypothesen die Rede ist, ist der Hypothesenbegriff im eben geschilderten Sinne zu verstehen.

**Hypothese 1a „Praktikabilität":**

**Zur Praktikabilität der HoNOS-D:** Aufgrund der Ergebnisse aus der Literatur kann erwartet werden, dass 9 von 12 Items **fehlende Werte** unter 10% aufweisen sollten. Für den Gesamtscore der HoNOS-D wird ebenfalls ein Anteil fehlender Werte unter 10% erwartet. Bei 3 der 12 Items (Item 8 „Andere Symptome", 11 „Probleme durch die Wohnbedingungen" und 12 „Probleme durch die Bedingungen im Beruf und Alltag") ist aufgrund der heterogenen Konzeption (Item 8) bzw. empirischen Hinweise hinsichtlich niedriger Interraterreliabilitäten (Item 11 und 12) ein prozentualer Anteil fehlender Werte über 10% zu erwarten.

Zur weiteren, ergänzenden Überprüfung der Praktikabilität der HoNOS-D sollen explorative strukturelle Einschätzungen von Therapeuten (n = 7), die die HoNOS-D eingesetzt haben, einbezogen werden.

**Hypothese 1b „Verteilungseigenschaften":**

Auf der Basis der Literatur kann angenommen werden, dass die **Verteilungseigenschaften** der Items 3 („Problematischer Alkoholkonsum") und 6 („Probleme in Zusammenhang mit Halluzinationen und Wahnvorstellungen") in der zu untersuchenden Stichprobe rechts schief verteilt sein werden (Deckeneffekte). Für alle anderen Items und den Gesamtscore der HoNOS-D werden gemäß den empirischen Befunden aus der Literatur moderate bis zufrieden stellende Verteilungseigenschaften (Schiefeindizes Wertebereich: -1 bis +1) erwartet.

**Hypothese 1c „Interkorrelation":**

Hinsichtlich der **Item-Interkorrelationen** werden vor dem Hintergrund der Instrumentenkonstruktion höchstens zwei signifikante Zusammenhänge mindestens mittlerer Effektstärke ($r > .40$) zwischen den Items der HoNOS-D erwartet. Es können aufgrund empirischer Befunde von Wing et al. (1996) signifikante Zusammenhänge mindestens mittlerer Effektstärke für

das Item 4 „Kognitive Probleme“ mit dem Item 10 „Alltägliche Aktivitäten“ und für das Item 2 „Selbstverletzung“ mit dem Item 7 „Gedrückte Stimmung“ erwartet werden.

**Hypothese 1d „Interraterreliabilität“:**

Bezüglich der **Interrater-Reliabilitäten** sollen aufgrund empirischer Befunde aus der Literatur für die Items 4 „Kognitive Probleme“, 8 „Andere verhaltensbezogene Probleme“, 9 „Probleme mit Beziehungen“, 10 „Probleme mit alltäglichen Aktivitäten“, 11 „Probleme durch die Wohnbedingungen“ und 12 „Probleme durch die Bedingungen im Beruf und Alltag“ niedrige Übereinstimmungen in den Koeffizienten (unter < .70) bzw. ein niedriger Prozentsatz der Einschätzungen genau der gleichen Kategorie zwischen zwei Beurteilern (unter 50%) erwartet werden. Für die übrigen Items und den Gesamtscore (Item 1 „Aggression“, Item 2 „Selbstverletzung“, Item 3 „Substanzmissbrauch“, Item 5 „Probleme mit körperlichen Erkrankungen“, Item 6 „Halluzinationen und Wahnvorstellungen“ und Item 7 „Gedrückte Stimmung“) können aufgrund der empirischen Befunde aus der Literatur (vgl. Abschnitt 6.3) befriedigende Werte der Koeffizienten zur Beurteilung der Interrater-Reliabilität (> .70) und befriedigende Prozentsätze der Einschätzungen genau der gleichen Kategorie (über 50%) erwartet werden.

**Hypothese 1e „Test-Retest-Reliabilität“:**

Aufgrund erster empirischer Befunde zur **Test-Retest-Reliabilität** werden befriedigende Werte der Interklassenkoeffizienten (>. 70) in 8 der 12 Items und dem Gesamtscore erwartet. Um den Einfluss von Behandlungseffekten auf die Test-Retest-Reliabilität zu kontrollieren, werden die entsprechenden Kennwerte an einer Substichprobe von Patienten bestimmt, die sich während der stationären Behandlung bezüglich zentraler Erfolgsmaße nicht verändert haben. Für 4 der 12 Items (Item 1 „Aggression“, 3 „Probleme mit Alkohol- oder Drogenkonsum“, 7 „Gedrückte Stimmung“ und Item 10 „Probleme mit alltäglichen Aktivitäten“) sind dagegen niedrigere Werte in den Intraklassenkorrelationskoeffizienten (< .70) zu erwarten.

Bezogen auf die Hauptfragestellung zur *spezifischen Eignung der HoNOS-D* lassen sich folgende Hypothesen ableiten:

**Hypothese 2a „Prädiktor HoNOS-D“:**

**Zur Eignung im Klassifikationssystem:** Auf der Basis der Ergebnisse des MH-CASC Projektes und des NZ-CAOS Projektes wird erwartet, dass der „Gesamtscore der HoNOS-D“ in einem **Regression-Tree-Verfahren** zur Vorhersage des Ressourcenverbrauchs (bzw. der Behandlungsdauer) während stationärer Rehabilitationsbehandlung extrahiert wird und somit einen Prädiktor innerhalb der Fallgruppenbildung von Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung darstellt.

Wenn die Hypothese 2a bestätigt werden kann, sollen folgende weitere Hypothesen getestet werden:

**Hypothese 2b „Bedeutsamkeit des Prädiktors HoNOS-D“:**

In einer **linearen Regression,** in der die **Behandlungsdauer** durch die Prädiktoren, die im Regression-Tree-Verfahren extrahiert werden, vorhergesagt werden soll, werden höhere Regressionsgewichte für den „Gesamtscore der HoNOS-D“ im Vergleich zu einem äquivalenten Fremdeinschätzungsscore („Psychische Summenskala der SF-8-F“) erwartet.

**Hypothese 2c „Position im Prädiktoren-Kern-Set“:**

Außerdem wird erwartet, dass der „Gesamtscore der HoNOS-D“ bei einem vorab definierten Set an Variablen, die sich aus der Regression-Tree-Analyse ergeben und zusätzlich in der Routineversorgung von Patienten mit psychischen Störungen vorhanden sind (**„Prädiktoren-Kern-Set“)** mittels Regression-Tree-Verfahren extrahiert und einen höheren Beitrag zur Varianzaufklärung liefert als ein vergleichbarer Fremdeinschätzungsscore („Psychische Summenskala der SF-8-F“).

Bezüglich der Fragestellung 3 werden aufgrund ihres explorativen Charakters keine expliziten Hypothesen formuliert.

# 9 Methode

## 9.1 Design der Untersuchung

Zur Beantwortung der genannten Fragestellungen wurde ein prospektives Design mit einem ersten Messzeitpunkt zu Beginn der stationären Behandlung gewählt. Die krankheits- und patientenbezogenen Kriterien wurden in der ersten Woche des stationären Aufenthaltes der Patienten per Fragebogen erhoben. Am Ende der Behandlung wurde die Kriteriumsvariable „Behandlungsdauer“ erfasst. Bei der Untersuchung handelt es sich damit um eine prospektive Feldstudie im Zwei-Punkt-Design zur Vorhersage des Ressourcenverbrauchs bzw. der Behandlungsdauer.

## 9.2 Rahmen der Untersuchung und Untersuchungssetting

Die Untersuchung wurde im Rahmen der Pilotphase des Qualitätssicherungsprogramms der Gesetzlichen Krankenkassen im Zeitraum von Februar 2003 bis Februar 2004 in 11 verschiedenen Rehabilitationskliniken des Indikationsbereiches „Psychische Störungen“ durchgeführt. Es handelt sich dabei um ein Qualitätssicherungsprojekt, welches von der Forschungsgruppe Qualitätsmanagement des Institutes für Medizinische Psychologie für die psychischen Indikationsbereiche angepasst und weiterentwickelt wurde. Dieses baut auf dem Programm für somatische Indikationsbereiche auf, das von der Abteilung „Qualitätsmanagement und Sozialmedizin am Universitätsklinikum Freiburg i. Br. (AQMS)“ in Abstimmung mit dem Arbeitskreis „Qualitätssicherung in der Rehabilitation“ des VdAK/AEV und der Fachebene der Spitzenverbände der Gesetzlichen Krankenkassen erarbeitet wurde.

Die Initiierung und Planung sowie die Organisation der Durchführung der Studie „Pilotphase des Qualitätssicherungsprogramms der Gesetzlichen Krankenkassen“ (QS-Reha) erfolgte durch die Projektleiter S. Kawski, H. Schulz, U. Koch sowie durch T. Harfst (Institut und Poliklinik für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf).

Das Qualitätssicherungsprogramm der Gesetzlichen Krankenkassen bezieht sich auf die Erhebung von fünf Komponenten: 1) Erfassung von Struktur-, 2) Prozess- und 3) Ergebnisqualität und 4) Erfassung von Patienten- und 5) Mitarbeiterzufriedenheit (Kawski & Koch, 2002). In

der vorliegenden Untersuchung wurden nur Daten aus der 3. Komponenten, der Messung der Ergebnisqualität verwendet.

Um einen Einblick in das für die Untersuchung gewählte Setting zu gewährleisten, sollen nun nachfolgend die 11 Kliniken in ihren strukturellen und prozessualen Komponenten beschrieben werden (Kawski, 2004).

10 der 11 Kliniken befinden sich entweder in privater Trägerschaft oder sind als eigenständige Gesellschaft (GmbH) organisiert. Alle Kliniken werden von den gesetzlichen Krankenkassen und der gesetzlichen Rentenversicherung belegt. Andere Kostenträger sind die Privatkrankenkassen und Beihilfestellen. Die Bettenanzahl der Klinken variiert mit einer Spanne von ca. 80 bis 250. Unter den Kliniken befindet sich eine spezialisierte Einrichtung für die Behandlung essgestörter Patienten. Die Behandlungsdauer in den Kliniken beträgt im Durchschnitt zwischen 4 bis 10 Wochen. Alle Kliniken zeichnen sich dadurch aus, dass sie mindestens 2 der 3 geforderten Richtlinienverfahren (kognitiv-verhaltenstherapeutisch, tiefenpsychologisch fundiert oder analytische fundierte Therapie) vorhalten (Kawski, 2004).

Es kann also festgehalten werden, dass es sich bei den elf Kliniken um eine weitgehend repräsentative Bandbreite von Rehabilitationseinrichtungen handelt, die an der stationären psychotherapeutischen Versorgung beteiligt sind.

### 9.3 Durchführung der Untersuchung

Für die Pilotphase des QS-Reha®-Projektes war mit den beteiligten Kliniken und den Kostenträgern anvisiert worden, eine konsekutive Stichprobe von jeweils n = 200 Patienten pro Klinik zu erheben, deren stationäre Rehabilitationsbehandlung durch die Gesetzlichen Krankenkassen finanziert werden. Dabei zeigte sich jedoch relativ zeitnah nach Erhebungsbeginn, dass diese Absprache zu einer deutlichen längeren Rekrutierungsphase geführt hätte, so dass die Vereinbarung noch einmal dahingehend verändert wurde, auch Patienten einzuschließen, deren Behandlung durch die Gesetzliche Rentenversicherung finanziert wurden.

Die Datenerhebung begann in den einzelnen Kliniken im Zeitraum vom 25.02.03 bis 02.05.2003, die letzten Patienten wurden vom 30.06.03 bis 08.11.03 in die Studie aufgenommen. Die Datenerhebungen zum Entlassungszeitpunkt waren je nach Klinik vom 30.06.03 bis

08.11.03 abgeschlossen, die letzten Fremdeinschätzungsbögen wurden am 18.02.04 an das auswertende Institut übersandt.

Zur Organisation der Datenerhebung wurde in den jeweiligen Kliniken ein Studienkoordinator als Ansprechpartner benannt. Die Studienkoordinatoren erhielten ein umfangreiches Manual, worin die einzelnen Erhebungsunterlagen, der Ablauf der Erhebungen und der Versand der Erhebungsunterlagen ausführlich erläutert wurden und dokumentiert waren.

Bei der Aufnahmeuntersuchung in den Kliniken wurde den Patienten zu Beginn (bzw. zum Ende) ihrer stationären psychosomatischen/psychotherapeutischen Rehabilitationsbehandlung ein Fragebogen zur Ergebnisqualität ausgehändigt (siehe Anhang 16.2.3). Auch die Therapeuten erhielten zu Beginn (bzw. zum Ende) der Behandlung einen Fragebogen, der zentrale Outcomeparameter enthielt (den sog. Arztbogen oder auch Fragebogen zur Ergebnisqualität, siehe Anhang 16.2.1 bzw. 16.2.2). Der Beurteilungszeitraum wurde auf die jeweils letzten 7 Tage festgelegt. Für Patienten, die aus verschiedenen Gründen nicht an der Untersuchung teilnehmen wollten, füllten die Therapeuten zusätzlich einen Fragebogen, den sog. Drop-Out-Bogen, aus. Patienten, die der deutschen Sprache nicht mächtig waren, wurden von der Untersuchung ausgeschlossen.

Ebenfalls zu Beginn der Behandlung wurde den Patienten eine Einverständniserklärung zur Teilnahme an der Studie ausgehändigt. Darin erklärten sie sich auch zur Weitergabe ihrer Daten in anonymisierter Form an das Institut und Poliklinik für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf und darüber hinaus zur katamnestischen Befragung sechs Monate nach ihrer stationären psychosomatischen/psychotherapeutischen Rehabilitationsbehandlung bereit. Lag die Einverständniserklärung vom Patienten vor, so wurden die Fragebögen zusammen mit den Therapeutenbögen an das Institut und Poliklinik für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf gesandt. Durch die Vergabe von Zahlencodes für jeden Patienten, die auf den patienten- und therapeutenseitig ausgefüllten Fragebögen aufgedruckt waren, wurden die Anonymität der Patienten und die korrekte Zuordnung der Fragebögen gewährleistet.

### 9.4 Stichprobenbeschreibung und Repräsentativitätsanalyse

Im Zeitraum von Februar 2003 bis Februar 2004 wurden insgesamt N = 2.386 Patienten um ihre Teilnahme an der Untersuchung gebeten. Von diesen nahmen insgesamt N = 468 (19%) Patienten aus verschiedenen Gründen (z.B. Therapieabbruch oder Verweigerung) nicht an der Studie teil (= Nonresponder). N = 243 (10%) Patienten konnten z.B. nicht berücksichtigt werden, da keine vollständigen Datensätze zu Beginn und am Ende der Behandlung vorhanden waren (= Partial-Nonresponder). Damit lag eine Gesamtstichprobe von N = 1.918 Patienten zur Überprüfung der Fragestellungen vor.

Für 5 der 11 Kliniken bzw. 176 von insgesamt 225 Patienten lagen Angaben über die Gründe einer Nicht-Teilnahme am Qualitätssicherungsprogramm vor. Tabelle 10 bietet dazu einen zusammenfassenden Überblick.

Tabelle 10: Gründe für Nicht-Teilnahme (Angaben aus 5 Kliniken)

| Gründe | Häufigkeit | Prozent |
|---|---|---|
| Verweigerung | 37 | 21 |
| Vorzeitige Abreise | 15 | 8 |
| Verlegung | 3 | 2 |
| Therapieabbruch | 12 | 7 |
| Anderes | 109 | 62 |
| Gesamt | 176 | 100 |

Von den 176 Nonrespondern mit vorliegenden Angaben zum Grund der Nicht-Teilnahme verweigerten 21% die Teilnahme, weitere 8% reisten vorzeitig ab und 7% brachen die stationäre Behandlung ab. Für den größten Teil der Nonresponder (62%) lagen andere Gründe als die vorgegebenen Kategorien „Verweigerung“, „Verlegung“ oder „Therapieabbruch“ zu. Genannt wurden hier u.a. mangelnde Motivation oder Widerruf der Teilnahmebereitschaft im Nachhinein.

Um die Repräsentativität der vorliegenden Untersuchungsstichprobe zu überprüfen, wurden Analysen zum Vergleich von Drop-Out-Patienten und den Studienteilnehmern vorgenommen.

Je nach Skalenniveau der zu überprüfenden Merkmale wurden t-Tests, chi²-Tests und Rangtests durchgeführt.

### 9.4.1 Drop-Out-Analyse: Auswertung relevanter Parameter aus dem Arztbogen

Um zu überprüfen, ob zwischen den n = 1918 Patienten der Untersuchungsstichprobe und den n = 468 „Drop-Out-Patienten“ Unterschiede in relevanten sozialmedizinischen und klinischen Parametern bestehen, wurde ein Vergleich hinsichtlich der folgenden Variablen aus dem Arztbogen (siehe Anhang 16.2.1) durchgeführt: „Diagnosegruppen“, „Chronifizierung der Erkrankung in Jahren“, „Motivation zur Rehabilitationsbehandlung“, „Gesamtwert im HoNOS-D“ und die „Gesamtwerte“ im Instrument zur „Fremdeinschätzung der gesundheitsbezogenen Lebensqualität für die Psychische und Somatische Summenskala“ („Short-Form-8-Fremdeinschätzung, SF-8-F“). Die Tabelle 11, Tabelle 12 und Tabelle 13 geben einen Überblick über die Ergebnisse der Analysen.

Tabelle 11: Unterschiede zwischen den Studienpatienten und den Drop-Out-Patienten hinsichtlich der Diagnosegruppen „Depressive Störungen“, „Angststörungen“ und „Somatoforme Störungen“ (chi²-Test)

| **Diagnose** | **Studienpatienten** | | **Drop-Out-Patienten** | | **Signifikanztests, Effektstärken** |
|---|---|---|---|---|---|
| | *n* | % | *n* | % | |
| | **1918** | **100** | **468** | **100** | |
| Depressive Störung | 1004 | 52,3 | 232 | 49,6 | $\chi^2 = 1,16$; $p = .28$; $phi = .02$ |
| Angststörung | 321 | 16,7 | 85 | 18,2 | $\chi^2 = 0,54$; $p = .46$ $phi = .02$ |
| Somatoforme Störung | 263 | 13,7 | 67 | 14,3 | $\chi^2 = 0,16$; $p = .73$ $phi = .01$ |

Tabelle 12: Unterschiede zwischen den Studienpatienten und den Drop-Out-Patienten hinsichtlich der Variable Chronifizierung in Jahren (Mann-Whitney-U-Test)

| Variable | Studienpatienten (n = 1823) | | Drop-Out-Patienten (n = 413) | | Signifikanztests, Effektstärken |
|---|---|---|---|---|---|
| | *Median* | *Mittlerer Rang* | *Median* | *Mittlerer Rang* | |
| **Chronifizierung** | 4 | 1110 | 4 | 1152 | *U-Wert* = 362549; *p* = .23; *Kendall-Tau-c* = .02 |

**Anmerkung:** Chronifizierung als Rangvariable (Rang 1 = Akutereignis, Rang 2 = 1 Jahr, Rang 3 = 1-2 Jahre, Rang 4 = 3-5 Jahre, Rang 5 = 5-10 Jahre und Rang 6 = mehr als 10 Jahre). Kendall´s Tau-c als Assoziationsmaß für die Effektstärke zwischen einem nominal- und ordinalskalierten Merkmal.

Tabelle 13: Unterschiede zwischen Studienpatienten und Drop-Out-Patienten hinsichtlich der Rehabilitationsmotivation, der Schweregradeinschätzung (HoNOS-Gesamtwert) und Fremdeinschätzungen der Lebensqualität (SF-8-F psychische und somatische Summenskala) (t-Test)

| Variable | Studienpatienten | | Drop-Out-Patienten | | Signifikanztests, Effektstärken[a] |
|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | |
| **Reha-Motivation**[d] | 2,57 | 1,15 | 3,06 | 1,33 | $t_{(df = 2295)} = 6{,}96^b$; $p \leq .001$; d = .40 |
| **HoNOS-Gesamtscore** | 1,38 | 0,47 | 1,46 | 0,59 | $t_{(df = 2294)} = 2{,}90^b$; $p \leq .01$; d = .18 |
| **Psychische Summenskala (SF-8-F)** | 2,59 | 0,64 | 2,63 | 0,70 | $t_{(df = 2334)} = 0{,}91^c$; $p = .36$; d = .05 |
| **Somatische Summenskala (SF-8-F)** | 1,79 | 0,81 | 1,89 | 0,83 | $t_{(df = 2335)} = 2{,}18^b$; $p = .03$; d = .11 |

**Anmerkung:**
[a] Signifikanztests und Effektmaße bezogen auf den Vergleich Untersuchungsteilnehmer und Drop-Out-Patienten
[b] t-Test bei inhomogenen Varianzen (Levene-Test)
[c] t-Test bei homogenen Varianzen (Levene-Test)
[d] Scoring der Reha-Motivation: 1 = sehr hohe Motivation bis 6 = keine Motivation

Zwischen den Studienpatienten und den Drop-Out-Patienten bestehen zu Beginn der Behandlung signifikante Unterschiede mit kleiner bzw. annähernd kleiner Effektstärke in der „Motivation zur Rehabilitationsbehandlung“ und im „Gesamtwert der HoNOS-D“ sowie mit weniger als einer kleinen Effektstärke in der „Somatischen Summenskala des SF-8-F“: die Drop-Out-Patienten werden als weniger motiviert und sowohl im HoNOS-D als auch im SF-8-F als höher belastet eingeschätzt.

Im Folgenden soll zur zentralen Drop-Out-Analyse eine ergänzende Subgruppenanalyse zwischen den Studienpatienten und denjenigen Patienten, zu denen sowohl zu Beginn als auch am Ende der Behandlung keine vollständigen Datensätze vorhanden waren (= Partial-Nonresponder), vorgenommen und deren Ergebnisse vorgestellt werden.

**Partial-Nonresponder: Auswertung relevanter Parameter aus dem Arzt- und Patientenbogen**

Eine weitere Subgruppenanalyse bezog sich auf den Vergleich der Partial-Nonresponder (n = 243, mit den Studienpatienten (n = 1918) in relevanten Parametern des Patientenbogens. Als Partial-Nonresponder wurden diejenigen Patienten definiert, von denen sowohl zu Beginn als auch am Ende der Behandlung keine vollständig auswertbaren Datensätze vorlagen.

Da von den Partial-Nonrespondern auch patientenseitig eingeschätzte Angaben vorhanden waren, wurden die Studienteilnehmer und die Partial-Nonresponder in den folgenden soziodemographischen und klinischen Merkmalen verglichen: „Alter", „Geschlecht", „Partnersituation", „Schulabschluss", „Erwerbssituation", „Rentenantrag", „GSI der SCL-14" (Global Severity Index der Symptom-Check-List-14, Harfst et al., 2002), „Gesamtwert der ADS-K" (Allgemeine Depressionsskala-Kurzversion, Hautzinger & Bailer, 2002) und „Gesamtwerte der patientenseitigen Erfassung gesundheitsbezogener Lebensqualität" (Short-Form-8-Selbsteinschätzung, SF-8-S, Ware et al., 2000). Die folgenden Tabellen 14 und 15 vermitteln einen Überblick über die Ergebnisse des Vergleiches zwischen Partial-Nonrespondern und Studienpatienten. Der Fragebogen für die Patienten zu Beginn der stationären psychosomatischen/psychotherapeutischen Rehabilitationsmaßnahme befindet sich im Anhang 16.2.3. Auf eine Darstellung des patientenseitigen Fragebogens zur Messung der Ergebnisqualität zum Ende der Behandlung wurde aus Platzgründen verzichtet.

Zwischen den Studienpatienten und den Partial-Nonrespondern (n = 243) bestehen signifikante Unterschiede mit annähernd kleiner bzw. weniger als einer kleinen Effektstärke in der Erwerbssituation und im Rentenantrag (siehe Tabelle 14). Die Studienteilnehmer sind signifikant häufiger berufstätig (49,4% vs. 42,5%), dagegen sind die Partial-Nonresponder signifikant häufiger als Hausfrau oder Hausmann beschäftigt (15,8% vs. 9,1%, siehe Tabelle 14). Weiterhin haben die Partial-Nonresponder signifikant häufiger als die Studienpatienten einen

Rentenantrag gestellt (9,1% vs. 5,1%). In den untersuchten klinischen Merkmalen unterscheiden sich die beiden Gruppen nicht signifikant voneinander (siehe Tabelle 15).

Tabelle 14: Unterschiede zwischen Studienpatienten und den Partial-Nonrespondern hinsichtlich soziodemographischer Merkmale (chi²-Test)

| Merkmale | Studienpatienten | | Partial-Nonresponder | | Signifikanztests, Effektstärken |
|---|---|---|---|---|---|
| | *n* | % | *n* | % | |
| **Geschlecht** | **1841** | **100** | **159** | **100** | $\chi^2 = 0,94$<br>$p = .33$<br>*phi* = .02 |
| männlich | 1417 | 77,0 | 117 | 73,6 | |
| weiblich | 424 | 23,0 | 42 | 26,4 | |
| **Partnersituation** | **1639** | **100** | **115** | **100** | $\chi^2 = 0,87$<br>$p = .83$<br>*phi* = .02 |
| <1 Jahr keinen Partner | 258 | 15,7 | 16 | 13,9 | |
| >1 Jahr keinen Partner | 456 | 27,8 | 29 | 25,2 | |
| fester Ehepartner | 664 | 40,5 | 50 | 43,5 | |
| fester Nicht-Ehepartner | 261 | 15,9 | 20 | 17,7 | |
| **Schulabschluss** | **1818** | **100** | **128** | **100** | $\chi^2 = 1,74$<br>$p = .63$<br>*phi* = .03 |
| Fachabitur/Abitur | 513 | 28,2 | 30 | 23,4 | |
| Realschulabschluss | 569 | 31,3 | 45 | 35,2 | |
| Hauptschulabschluss | 551 | 30,3 | 41 | 32,0 | |
| ohne Abschluss, sonstiges | 185 | 10,2 | 12 | 9,4 | |
| **Erwerbssituation** | **1733** | **100** | **120** | **100** | $\chi^2 = 14,14$<br>$p = .02$<br>*phi* = .09 |
| berufstätig | 856 | 49,4 | 51 | 42,5 | |
| arbeitslos | 232 | 13,4 | 13 | 10,8 | |
| Hausfrau/-mann | 158 | 9,1 | 19 | 15,8 | |
| Altersrente | 98 | 5,7 | 5 | 4,2 | |
| EU-Rente | 109 | 6,3 | 15 | 12,5 | |
| sonstiges | 280 | 16,2 | 17 | 14,2 | |
| **Rentenantrag** | **1779** | **100** | **198** | **100** | $\chi^2 = 5,4$<br>$p = .02$<br>*phi* = .05 |
| ja | 91 | 5,1 | 18 | 9,1 | |
| nein | 1688 | 94,9 | 180 | 90,9 | |

Tabelle 15: Unterschiede zwischen Studienpatienten und den Partial-Nonrespondern hinsichtlich der klinischen Merkmale (t-Test)

| **Merkmale** | **Studienpatienten** | | **Partial-Nonresponder** | | **Signifikanztests, Effektstärken**[a] |
|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | |
| **Alter** | 42,3 | 14,0 | 40,42 | 14,33 | $t_{(df = 1997)} = 1,64$[b]; *p* = .10; d = .14 |
| **GSI der SCL-14** | 1,27 | 0,74 | 1,33 | 0,80 | $t_{(df = 1999)} = 0,88$[b]; *p* = .381; d = .08 |
| **ADS-K-Gesamtwert** | 22,11 | 9,67 | 23,05 | 10,2 | $t_{(df = 1941)} = 1,10$[b]; *p* = .29; d = .10 |
| **Psychische Summenskala (SF-8)** | 2,26 | 0,89 | 2,38 | 0,90 | $t_{(df = 2024)} = 1,47$[b]; *p* = .14; d = .13 |
| **Somatische Summenskala (SF-8)** | 1,83 | 0,95 | 1,91 | 0,94 | $t_{(df = 2036)} = 0,94$[b]; *p* = .35; d = .08 |

[a] Signifikanztests und Effektmaße bezogen auf den Vergleich QS-Rehateilnehmer und Partial-Nonresponder (Selbsteinschätzung)
[b] t-Test bei homogenen Varianzen (Levene-Test)

## Zusammenfassende Bewertung

Die Drop-Out-Patienten (Nonresponder und Partial-Nonresponder) weisen nach Einschätzung der Therapeuten eine geringere Motivation zur stationären Behandlung auf als die Studienteilnehmer. Weiterhin sind sie nach Einschätzung der Behandler symptomatisch stärker belastet (SF-8-F „Somatische Summenskala“; HoNOS-D). Ein weiteres Unterscheidungsmerkmal betrifft die Erwerbstätigkeit und die Rentenantragstellung. Die Partial-Nonresponder sind signifikant weniger berufstätig und haben signifikant häufiger einen Rentenantrag gestellt als die Studienteilnehmer.

Da die gefundenen Unterschiede zwischen den Studienpatienten und den Drop-Out Patienten durchgängig höchstens kleine Effektstärken annehmen, kann für die vorliegende Untersuchungsstichprobe eine weitestgehende Repräsentativität für die Patienten der Kliniken angenommen werden.

### *9.4.2 Stichprobenbeschreibung*

Die konsekutiv gewonnene Untersuchungsstichprobe umfasste N = 1.918 Patienten. Obgleich wesentliche Stichprobencharakteristika den oben aufgeführten Tabellen zu entnehmen sind, soll an dieser Stelle noch einmal ein Gesamteindruck über die Untersuchungsstichprobe ver-

mittelt werden. Die Verteilung weiterer relevanter soziodemographischer und klinischer Merkmale der Stichprobe ist in Tabelle 16 dargestellt.

Darüber hinaus sind die Patienten der Stichprobe dadurch gekennzeichnet, dass 46% bei Aufnahme nicht arbeitsfähig und 31% zu Beginn der stationären Behandlung krankgeschrieben waren. Die durchschnittliche Behandlungsdauer der 1918 Patienten betrug 47.5 Tage (SD = 17 Tage). Die Analyse ohne die Patienten der Spezialklinik für Essstörungen (n = 1695) zeigte eine mittlere Liegezeit von 43.8 Tagen auf (SD = 14 Tage).

Tabelle 16: Verteilung von soziodemographischen und klinischen Charakteristika (N = 1918)

| | **Stichprobe (N = 1918)** | |
|---|---|---|
| **Merkmal** | *Anzahl* | *Prozent (%)* |
| **Geschlecht** | | |
| Anteil Frauen | 1417 | 74 |
| **Alter (M = 42, SD = 14)** | | |
| < 29 Jahre | 373 | 20 |
| 30-39 Jahre | 360 | 19 |
| 40-49 Jahre | 516 | 27 |
| 50-59 Jahre | 390 | 20 |
| > 60 Jahre | 201 | 11 |
| **Schulbildung** | | |
| Hauptschule/Volksschule | 551 | 29 |
| Realschule/mittlere Reife | 569 | 30 |
| Abitur/Fachhochschulreife | 513 | 27 |
| Sonstiges | 185 | 10 |
| **Hauptdiagnose (ICD-10)** | | |
| Depressive Störung (F32-F34) | 728 | 38 |
| Essstörung (F50) | 289 | 15 |
| Belastungs-/Anpassungsstörung (F43) | 253 | 13 |
| Angststörung (F40-F41) | 173 | 9 |
| Somatoforme Störung (F45) | 86 | 5 |
| Sonstige Störungen | 386 | 20 |

**Anmerkungen:** Differenzen zur Gesamtstichprobe und zu 100 Prozent sind auf fehlende Werte in den einzelnen Variablen zurückzuführen.

Zur Beurteilung der Interrater-Reliabilitäten der HoNOS-D lagen zwei ergänzend erhobene Stichproben - einen aus der Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie und zum anderen aus stationärer Rehabilitationsbehandlung Suchtkranker – vor. Die Beurteilerpaare wurden zumeist durch den behandelnden Therapeuten und den Oberarzt der Einrichtung gebildet. Die hier berichteten Einschätzungen wurden **vor** der Durchführung ei-

ner Schulung im HoNOS-D erhoben, so dass es sich also um nicht-trainierte Rater handelt. Da die Datenerhebung noch nicht abgeschlossen ist, liegen auch noch keine umfassenden Angaben zur Stichprobenzusammensetzung vor, weswegen die Einrichtungen und das dort übliche Klientel im Folgenden nur grob skizziert werden sollen.

Das Klinikkonzept der stationären Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie ist durch die Behandlung mit vorwiegend psychodynamischen Ansätzen mit verhaltenstherapeutischen Elementen gekennzeichnet. Die Patienten sind zu einem hohen Anteil Frauen (ca. 75%), im Durchschnitt 40 Jahre alt und weisen als häufigste Erstdiagnose eine depressive Störung, gefolgt von Angststörungen und Anpassungsstörungen auf (Sammet et al., in press).

Die Patienten der stationären Rehabilitationseinrichtung für Suchtkranke sind demgegenüber dadurch gekennzeichnet, dass der Anteil der Männer bei 73% liegt, ca. 68% der Patienten älter als 40 Jahre sind und die Hauptsuchtmitteldiagnose zumeist Störungen durch Alkohol lautet (Zobel et al., 2004).

## 9.5 Operationalisierung der Prädiktoren des Ressourcenverbrauchs

Zur Prüfung der speziellen Eignung der HoNOS-D innerhalb der Fallgruppenbildung sollen größtenteils bereits etablierte Instrumente und standardisierte Items eingesetzt werden. Wie bereits unter Abschnitt 5.2 erwähnt wurde, sollten dabei Fremd- *und* Selbstbeurteilungen gleichermaßen einbezogen werden.

### *9.5.1 Selbsteinschätzung zu Beginn der Behandlung*

Im Folgenden soll zunächst die Operationalisierung der Kriterien zur Selbsteinschätzung dargestellt werden. Zur Überprüfung der Skalenstruktur der verwendeten Instrumente werden anhand der vorliegenden Daten konfirmatorische Faktorenanalysen berechnet und Itemanalysen durchgeführt, welche anschließend berichtet werden.

#### 9.5.1.1 Verwendete Selbsteinschätzungsinstrumente und Items

Tabelle 17 gibt einen Überblick über die in der vorliegenden Untersuchung verwendeten Instrumente und Items. Eine Überprüfung von bereits gut etablierten Selbsteinschätzungsinstrumenten zur Erfassung der Symptombelastung von Patienten mit psychischen Störungen wurde bereits allgemein in Kapitel 5 und speziell in Abschnitt 5.2 vorliegender Arbeit vorgenommen. Dabei wurde hervorgehoben, dass die Kurzversionen der SCL (SCL-14, Harfst et al., 2002), der SF-8, Ware et al., 2000) und des IIP (IIP-25, Harfst et al., 2004b) *zentrale* Selbsteinschätzungsinstrumente bei der Bildung von Fallgruppen für Patienten mit psychischen Störungen darstellen. Zur differenzierten Erhebung depressiver Symptomatik soll außerdem eine in der Kurzversion vorliegende Form der Allgemeinen Depressionsskala (ADS-K, [11] Hautzinger & Bailer, 1993) herangezogen werden.

Weitere als ressourcenverbrauchsrelevant identifizierte Kriterien der krankheitsbezogenen Dimension (vgl. Kap. 3, Abschnitt 3.3 und Kap. 4, Abschnitt 4.3) beziehen sich auf die Erhebung vorheriger Behandlungen (psychiatrische und psychotherapeutische, ambulante und stationäre Therapien) und auf medikamentöse Behandlung zum Zeitpunkt der Aufnahme. Diese Kriterien werden über standardisierte Items der PsyBado operationalisiert (Heuft & Senf, 1998) (siehe Tabelle 17).

Die Kriterien der patientenbezogenen Dimension werden über standardisierte Items zur Erhebung soziodemografischer Aspekte operationalisiert (siehe Tabelle 17). Zusätzlich sollen standardisierte Items mit spezifischen rehabilitativen Aspekten, wie z.B. Antrag auf Erwerbsminderungsrente oder Arbeitsunfähigkeitszeiten in der Fallgruppenbildung verwendet werden.

---

[11] In einer Überprüfung zeigte sich die ADS-K hinsichtlich der Güte psychometrischer Eigenschaften und der Praktikabilität anderen Instrumenten zur Erfassung depressiver Symptomschwere als überlegen (z.B. der HADS (Herrmann et al., 1995) und dem BDI (Hautzinger et al., 1995)). Zur ausführlicheren Darstellung der psychometrischen Kennwerte und der Bewertung siehe Abschlussbericht des Projektes „Entwicklung eines Klassifikationssystems für Patienten mit psychischen Störungen in stationärer Rehabilitationsbehandlung“ (Andreas et al., 2005).

Tabelle 17: Prädiktoren der krankheits- und patientenbezogenen Dimension und deren Operationalisierungen (Selbsteinschätzung)

| Dimension | Prädiktoren | Operationalisierung |
|---|---|---|
| **Krankheitsbezogen** | Symptombelastung | • Gesamtbelastungsindex GSI der SCL-14 (Harfst et al., 2002)<br>• Gesamtscore der ADS-K (Hautzinger & Bailer, 1993)<br>• Psychische und Somatische Summenskala der SF-8-S (Ware et al., 2000)<br>• Subskalen des IIP-25 (Harfst et al., 2004b) |
| | Vorherige Behandlungen | • Anzahl ambulanter und stationärer psychotherapeutischer Vorbehandlungen (Heuft & Senf, 1998) |
| | Belastung durch medikamentöse Therapie | • Einnahme von Schmerz- und Schlafmitteln (Heuft & Senf, 1998)<br>• Einnahme von Antidepressiva (Heuft & Senf, 1998) |
| **Patientenbezogen** | Soziodemografische Charakteristika | • Alter<br>• Geschlecht<br>• BMI<br>• Partnersituation<br>• Schulabschluss<br>• Berufsabschluss<br>• aktuelle berufliche Situation<br>• derzeit arbeitsunfähig/krankgeschrieben<br>• Antrag auf Erwerbsminderungsrente<br>• Absicht, Antrag auf Erwerbsminderungsrente zu stellen |
| | Sozialmedizinische Merkmale | • Arbeitsunfähigkeitszeiten für Berufstätige und Nicht-Berufstätige |

#### 9.5.1.2 Überprüfung und Kennwerte der Selbsteinschätzungsinstrumente und standardisierten Items

Wie gerade dargelegt, sollen bereits standardisierte Selbsteinschätzungsinstrumente in der Untersuchung verwendet werden. Da es sich jedoch um relativ neu entwickelte Instrumente handelt, soll im Folgenden noch einmal eine Überprüfung der Skalenstruktur und der psychometrischen Kennwerte der Instrumente an der Untersuchungsstichprobe vorgenommen werden.

Nach einer kurzen Einführung zum Entwicklungshintergrund und ausgewählten psychometrischen Kennwerten der Instrumente werden die Ergebnisse der Überprüfung der Skalenstruktur mittels konfirmatorischer Faktorenanalysen vorgestellt.

*Allgemeine Erläuterungen zur konfirmatorischen Faktorenanalyse und zum methodischen Vorgehen*

Zur Überprüfung der Skalenstruktur z.B. der SCL-14 und ADS-K in der vorliegenden Untersuchungsstichprobe wurden konfirmatorische Faktorenanalysen (CFA) gerechnet (Programmpaket AMOS 4.0). 1996 von Ritz et al. weiterentwickelt, stellt die CFA einer Weiterentwicklung der Faktorenanalyse dar (zitiert nach Bühner, 2004). Die Zielsetzung der Methode besteht darin, latente Variablen zu erfassen und Indikatorvariablen kausal auf diese zurückzuführen. Gleichzeitig liefert sie damit eine Methode zur Überprüfung theoretisch bzw. empirisch fundierter Modelle hinsichtlich ihrer Modellgüte.

Zur Schätzung der Modellparameter stehen verschiedene Verfahren zur Verfügung (z.B. die Maximum-Likelihood (ML) oder die Generalized Least Squares (GLS) Methode), die von der Skalenqualität, der Verteilungsart und der Stichprobengröße abhängig sind. Eine Entscheidung für die ML-Methode sollte von der Voraussetzung abhängig sein, ob die Daten normalverteilt und intervallskaliert sind bzw. eine Stichprobengröße von $N > 100$ vorliegt. Verletzungen dieser Voraussetzung können zu überhöhten Schätzungen des chi²-Wertes und zu moderaten Unterschätzungen der Standardfehler der Parameterschätzungen führen (Bühner, 2004).

Bühner (2004) empfiehlt die ML-Methode als Standardverfahren zur Parameterschätzung ab Stichprobengrößen von $N > 100$ und multivariater Normalverteilung. Dieser Empfehlung folgend, wurde diese Methode bei den folgenden Berechnungen zugrunde gelegt.

Eine nächste Entscheidung betrifft die verschiedenen Gütekriterien oder auch Fit-Indizes, die zur Bewertung der mittels konfirmatorischer Faktorenanalysen ermittelten Modellstruktur zur Verfügung stehen. Dabei ist einer der zentralen Fit-Indizes der chi²-Wert. Dieser wird allerdings bei größeren Stichprobenumfängen sensitiver für Modellfehlspezifikationen und führt damit zu einer erhöhten Wahrscheinlichkeit der Verwerfung der Nullhypothese (H0: Das Modell passt zur Datenstruktur) (Bühner, 2004). Deswegen behilft man sich weiterer Fit-Indizes, die sich hinsichtlich ihrer Sensitivität gegenüber Modellfehlspezifikationen und hinsichtlich der Abhängigkeit vom Stichprobenumfang voneinander abgrenzen lassen. Dabei kann man zwischen absoluten und inkrementellen Fit-Indizes unterscheiden. Während absolute Fit-Indizes den Anteil der Varianzaufklärung an der Gesamtvarianz anzeigen, beziehen sich in-

krementelle Gütekriterien auf die proportionale Verbesserung der Anpassung gegenüber einem restriktiveren Modell.

In diesem Zusammenhang empfehlen Hu und Bentler (1999), bei der Anwendung der ML-Methode zusätzlich zum chi²-Wert zwei weitere Fit-Indizes zur Prüfung der Güte des Modells zu verwenden. Eine Empfehlung bezieht sich dabei u.a. auf die Verwendung der beiden absoluten Fit-Indizes, nämlich dem Standardized-Root-Mean-Square-Residual (SRMR) in Kombination mit dem Root-Mean-Square-Error-of-Approximation (RMSEA). Die empfohlenen Cut-off Werte der beiden Kennwerte sind: RMSEA $\leq$ .06 und SRMR $\leq$ .09 (oder .10), die bei dieser Zusammensetzung in der Summe zu den geringsten Typ I und Typ II Fehlerraten führen.

Nach Prüfung der Voraussetzungen von Normalverteilung und Intervallskalenniveau (vergleiche hierzu auch Abschnitt 9.5.1.2.5 in Kap. 9) wurde für die nun folgenden Berechnungen die ML-Methode gewählt. Wie dargelegt, birgt der chi²-Wert das Problem in sich, dass er bei größeren Stichprobenumfängen sensitiver für Modellspezifikationen wird. Deswegen wird der chi²-Wert nicht berichtet und stattdessen in Anlehnung an die Empfehlungen von Hu und Bentler (1999) die Fit-Indizes SRMR in Kombination mit RMSEA mitgeteilt.

Der Kennwert SRMR wird allerdings nur dann berechnet, wenn keine fehlenden Werte in den Instrumenten vorliegen. Deswegen wurde es notwendig vor der Berechnung einer konfirmatorischen Faktorenanalyse eine Ersetzung fehlender Werte vorzunehmen. Dazu wurde in Anlehnung an die Empfehlung von Wirtz (2004) auf den Expectation-Maximation-Algorithmus (EM-Algorithmus) zurückgegriffen. Die Voraussetzungen zur Anwendung des EM-Algorithmus (Stichprobenumfang $N \geq 100$ und Prozent fehlender Werte $\leq$ 30% in der Stichprobe pro Variable) können als erfüllt angesehen werden (vergleiche hierzu auch Tabelle 18 im Abschnitt 9.5.1.2.5, Kap. 9). Der EM-Algorithmus folgt hierbei dem Prinzip, dass die im Datensatz vorhandene Informationsstruktur gültig ist, die aus Mittelwerten, Standardabweichungen und Korrelationen besteht. Auf der Basis dieser Informationen werden die fehlenden Werte zunächst geschätzt und dann ersetzt. Anschließend werden dann wiederum die Informationsträger (Mittelwert, Standardabweichungen und Korrelationen) erneut berechnet. Liegen dann noch Abweichungen vom Originaldatensatz vor, so kann geschlussfolgert werden, dass die ursprüngliche Informationsstruktur nicht gültig sein konnte und die neu berechnete Informationsstruktur wird beibehalten. Auf deren Basis werden dann die Kennwerte er-

neut geschätzt und dann ersetzt. Dieses Vorgehen wird solange wiederholt, bis eine Struktur gewonnen wurde, die sich nach der Ersetzung nicht mehr verändert.

#### *9.5.1.2.1 Die Symptom-Check-List (SCL-14)*

Harfst et al. (2002) entwickelten die SCL-14 anhand einer Stichprobe von N = 6962 Patienten aus drei verschiedenen psychotherapeutischen Fachkliniken. Aus den Items der SCL-90-R (Franke, 1995) wurden mittels explorativer und konfirmatorischer Faktorenanalysen die 14 ladungsstärksten Items ausgewählt und 3 Skalen gebildet („Depressivität“, „Phobische Angst“ und „Somatisierung“).

Zu den psychometrischen Eigenschaften der SCL-14 lässt sich festhalten, dass sich die internen Konsistenzen der drei Skalen in einem zufrieden stellenden Bereich zwischen r = .80 (Somatisierung) und r = .87 (Depressivität) befinden. Die interne Konsistenz für den Gesamtbelastungswert (GSI-14) beträgt r = .87. Harfst et al. (2002) berichten weiterhin einen signifikant hohen Zusammenhang zwischen GSI-14 und dem GSI der SCL-90-R (r = .92). Auch die Prüfungen der Kriteriumsvalidität ergaben zufrieden stellende Werte: So konnten die drei Skalen zwischen spezifischen Diagnosegruppen („Agoraphobie mit bzw. ohne Panikstörung“, „Somatoforme Störungen“ und „Depressive Störungen“) gut differenzieren. Die so entwickelte Kurzform wurde anschließend an einer weiteren klinischen Stichprobe validiert und zeigte auch dort gute Ergebnisse.

In Abbildung 14 ist die zugrunde liegende und hier erneut überprüfte Skalenstruktur der drei Skalen der SCL-14 graphisch veranschaulicht. Es sind weiterhin die partiellen standardisierten Regressionsgewichte der einzelnen Items auf den Faktoren angegeben, die auch als Korrelationen interpretiert werden können. Mit der konfirmatorischen Faktorenanalyse konnte die postulierte Skalenstruktur der SCL-14 in der Untersuchungsstichprobe bestätigt werden. Die beiden ausgewählten Anpassungstests zeigen zufrieden stellende Ergebnisse. Die Koeffizienten sind kleiner als .09 (SRMR = .038) bzw. kleiner als .06 (RMSEA = .054).

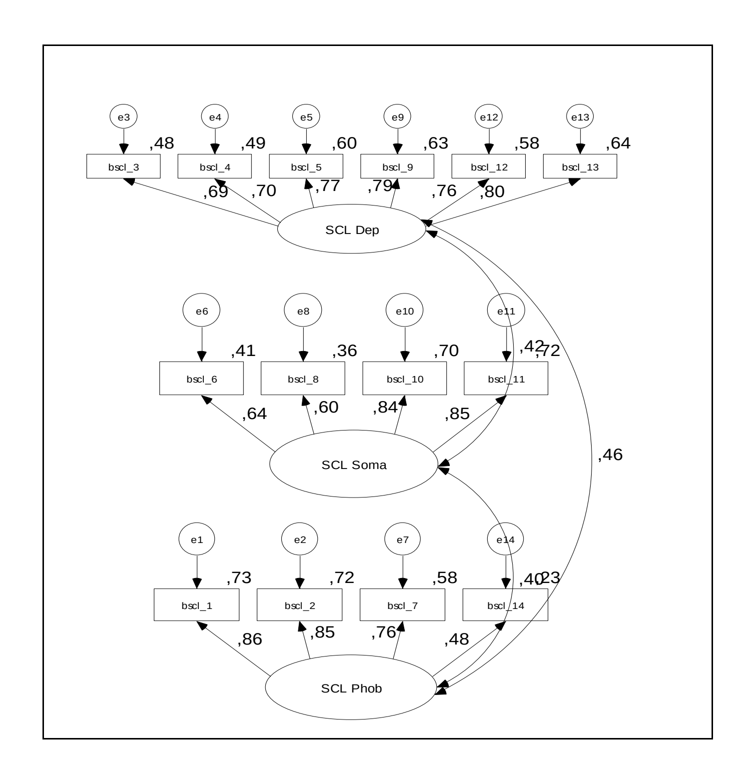



Abbildung 14: Postulierte und mittels konfirmatorischer Faktorenanalyse überprüfte Skalenstruktur der drei Skalen der SCL-14

#### *9.5.1.2.2 Die Short-Form-8 (SF-8-S)*

Eine Darstellung der Entwicklung der SF-8 und ihre psychometrischer Eigenschaften erfolgte bereits im Abschnitt 5.1.2 (Kap. 5) vorliegender Arbeit und kann dort entnommen werden.

Abbildung 15 veranschaulicht die Skalenstruktur der SF-8. Die partiellen standardisierten Regressionsgewichte weisen alle Werte über .60 auf und können damit als zufrieden stellend interpretiert werden. Die konfirmatorische Faktorenanalyse bestätigt größtenteils die zugrunde liegende Skalenstruktur der SF-8: Während der SRMR in einem akzeptablen Bereich liegt (SRMR = .07), befindet sich der RMSEA noch im akzeptablen Bereich, ist aber leicht erhöht (RMSEA = .11). Die Modellstruktur der zwei Skalen der SF-8-S kann somit als bestätigt angesehen werden.

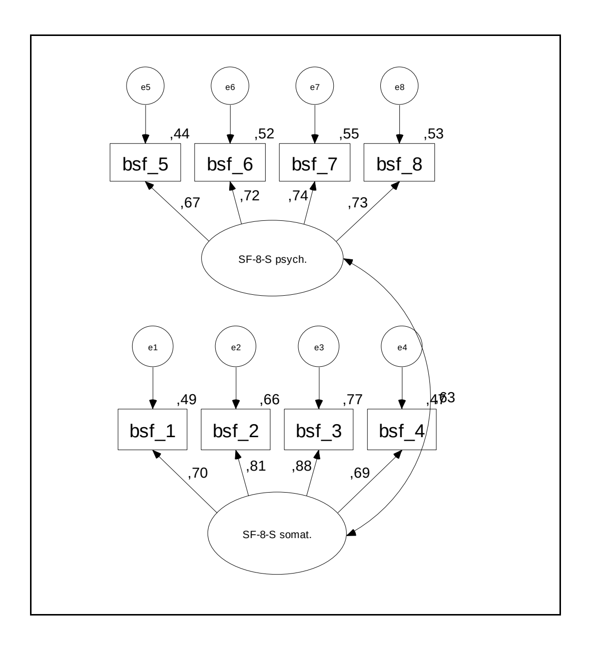



Abbildung 15: Postulierte und mittels konfirmatorischer Faktorenanalyse überprüfte Skalenstruktur der zwei Skalen der SF-8-S

*9.5.1.2.3 Die Allgemeine Depressionsskala (ADS-K)*

Bei der Allgemeinen Depressionsskala ADS (Hautzinger & Bailer, 2002) handelt es sich um die deutsche Übersetzung der Center for Epidemiological Studies Depression Scale (CES-D, Radloff, 1977), welche bereits 1977 für den Einsatz in großen epidemiologischen Studien entwickelt wurde. Die in der vorliegenden Untersuchung eingesetzte Kurzversion (ADS-K) enthält 15 Items, welche umfangreich depressive Symptomatik erfassen und auf einer Skala von 0 bis 3 („selten bis „die ganze Zeit“) beantwortet werden können. Als Index der depressiven Belastungsintensität steht ein Summenwert zur Verfügung.

Die psychometrischen Eigenschaften der ADS-K können als gut bewertet werden. So werden interne Konsistenzen über r = .90 berichtet. Auch die Korrelation zu anderen Selbstbeurteilungsmaßen depressiver Symptomatik (z.B. dem Beck Depressionsinventar BDI, Hautzinger et al., 1995) kann als gut bewertet werden (r = .72 und .94). Zusammenhänge zu Fremdbeurteilungsinstrumenten (wie z.B. der Hamilton Depressionsskala HAMD, Hamilton, 2002) werden mit Koeffizienten zwischen r = .49 und .86 berichtet. Eine von den Autoren gerechnete Faktorenanalyse legt die ausschließliche Verwendung des Gesamtsummenwertes der ADS-K nahe. Weiterhin stehen für die ADS-K Vergleichs- und Normwerte für verschiedene Altersgruppen und beide Geschlechter zur Verfügung.

Abbildung 16 veranschaulicht die geprüfte Modellstruktur der ADS-K der konfirmatorischen Faktorenanalyse. Auch hier konnte die postulierte Skalenstruktur zufrieden stellend durch die konfirmatorische Faktorenanalyse bestätigt werden (vgl. Abbildung 16). Die Koeffizienten der beiden Anpassungstests befinden sich in einem akzeptablen Bereich mit Werten von SRMR = .05 und RMSEA = .09).

##### *9.5.1.2.4 Das Inventar zur Erfassung Interpersoneller Probleme (IIP-25)*

Harfst et al. (2004b) entwickelten an einer Stichprobe von N = 1683 Patienten in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung mittels exploratorischer und konfirmatorischer Faktorenanalysen eine Kurzversion des Inventars zur Erfassung Interpersoneller Probleme (IIP-25). Auf der Basis der Ergebnisse einer explorativen Faktorenanalyse extrahierten Harfst et al. (2004) zunächst die ladungsstärksten 25 Items von insgesamt 64 Items des IIP (Horowitz et al., 1994). Die so gewonnene Faktorenstruktur überprüften sie dann an einer weiteren Stichprobe mittels konfirmatorischer Faktorenanalyse. Diese zeigte zufrieden stellende Ergebnisse (CFI = .84, SRMR = .07). Die internen Konsistenzen der 5-faktoriellen Lösung betrugen zwischen r = .76 und r = .84 und können als befriedigend bewertet werden. Hinsichtlich der Kriteriumsvalidität zeigten sich ebenfalls größtenteils befriedigende Ergebnisse. So wiesen z.B. Patienten mit einer sozialen Phobie signifikant höhere Werte auf der Subskala des IIP-25 „sozial-vermeidend“ auf.

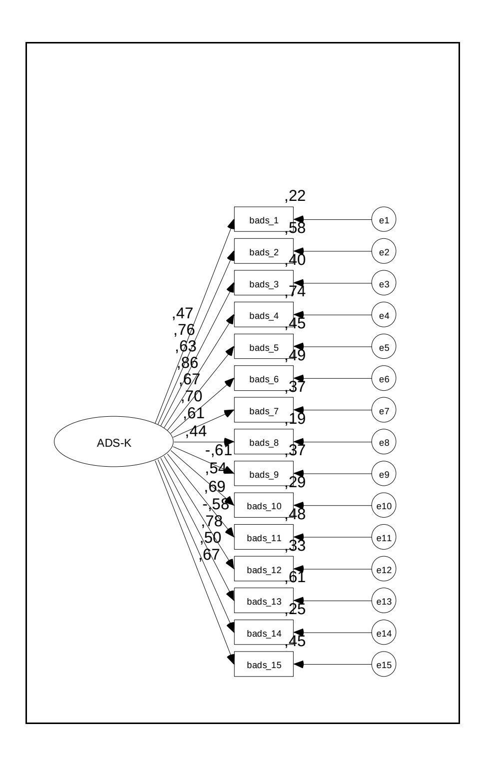



Abbildung 16: Postulierte und mittels konfirmatorischer Faktorenanalyse überprüfte Skalenstruktur der ADS-K

Abbildung 17 zeigt die zugrunde liegende Skalenstruktur des IIP-25. Die Ergebnisse der konfirmatorischen Faktorenanalyse bestätigen den postulierten Aufbau des Instrumentes in der Untersuchungsstichprobe. Die beiden Anpassungstests zeigen gute Ergebnisse. Der SRMR ist kleiner als .09 (SRMR = .07) und der RMSEA beträgt .06.

Nachdem die Skalen der eingesetzten Selbsteinschätzungsinstrumente faktorenanalytisch überprüft und in ihren Skalenstrukturen bestätigt werden konnten, sollen nachfolgend weitere psychometrische Kennwerte und Verteilungscharakteristika der standardisierten intervallskalierten Items berichtet werden.

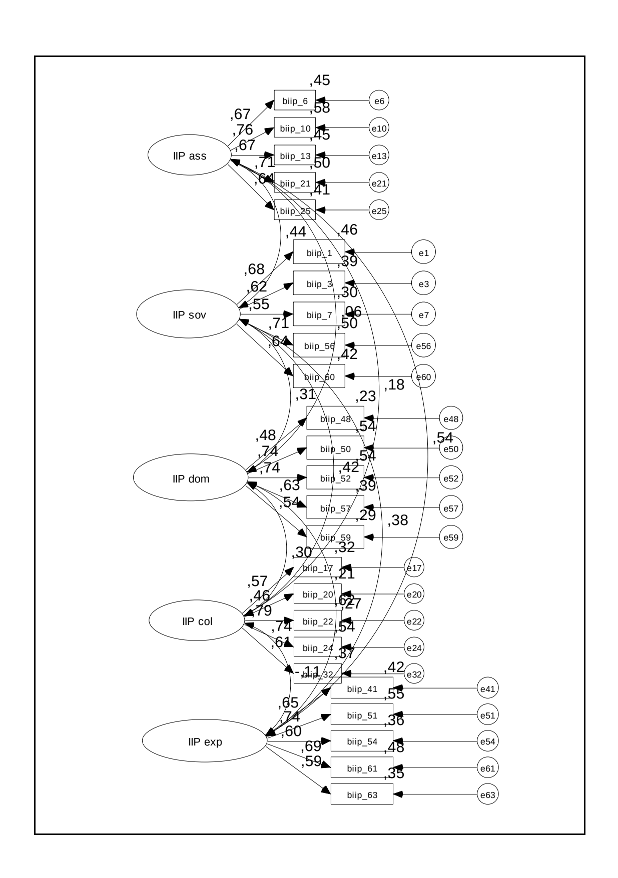



Abbildung 17: Postulierte und mittels konfirmatorischer Faktorenanalyse überprüfte Skalenstruktur des IIP-25

#### *9.5.1.2.5 Psychometrische Kennwerte der Skalen und standardisierten metrischen Items*

Tabelle 18 gibt einen Überblick über die psychometrischen Kennwerte der *Skalen* der eingesetzten *Selbsteinschätzungsinstrumente*.

Dabei ergab sich für die Konstruktion der Skalen („Psychische und Somatische Summenskala“) der **SF-8** zunächst das Problem, dass den Items unterschiedliche Antwortformate zugrunde lagen (vgl. auch hierzu ausführlich Abschnitt 5.1.2). Laut Testhandbuch wird hierbei empfohlen, jedes Item zu gewichten und durch einen aggregierten Score zu ersetzen (Ware et al., 2000). Dieses Vorgehen hat aber zur Folge, dass die Skalen nicht mehr berechnet werden können, sobald ein fehlender Wert vorliegt. Um die Anzahl gültiger Werte für die SF-8 zu erhöhen, wurden die Items zunächst, gemäß dem von Lang beschriebenen methodischen Vorgehen, auf gleiche Skalenbreite transformiert (Lang, 2003, S. 133). Anschließend wurden ungewichtete Mittelwerte für die somatische und psychische Summenskala der gesundheitsbezogenen Lebensqualität berechnet. Dabei mussten mindestens 3 von 4 Items der jeweiligen

Skala beantwortet sein. Das entspricht einem Kriterium von mindestens 75% gültiger Werte und folgt den Empfehlungen von Wirtz (2004)[12].

Eine deskriptive Analyse ergab für die beiden Subskalen der SF-8 **„Psychische Summenskala“** 1.6% und für die **„Somatische Summenskala“** 1.0% fehlende Werte in der Untersuchungsstichprobe. Die internen Konsistenzen betragen für die „Psychische Summenskala“ der SF-8 α = .80 und für die „Somatische Summenskala“ α = .84 und liegen damit innerhalb des Bereiches der im Testhandbuch berichteten Koeffizienten (siehe Tabelle 18).

Die deskriptive Analyse der **drei Subskalen** und des **GSI der SCL-14** ergab gegenüber der SF-8 eine leicht erhöhte Prozentzahl fehlender Werte (zwischen 2.2 und 2.7%). Die Subskala „Phobische Angst“ ist dabei erwartungsgemäß rechtsschief verteilt und weist auch den höchsten Wert in der Kurtosis auf. Cronbach´s alpha für die Subskalen und den Gesamtbelastungsindex GSI der SCL-14 nehmen Werte zwischen .82 und .89 an und liegen damit in dem von Harfst et al. (2002) berichteten Bereichen (siehe Tabelle 18).

Die deskriptive Analyse des **Gesamtwertes der ADS-K** zeigte gegenüber den drei Subskalen der SCL-14 wiederum einen leicht erhöhten prozentualen Anteil fehlender Werte (5.3%). Cronbach´s alpha beträgt für den Gesamtwert der ADS-K .83 und liegt etwas unterhalb des im Testhandbuch berichteten Kennwertes, aber dennoch in einem zufrieden stellenden Bereich (siehe Tabelle 18).

Auch bei der **Subskala „kühl/abweisend“ des IIP-25** liegt die Prozentzahl fehlender Werte mit 5.2% in einem für die ADS-K vergleichbaren Bereich. Cronbachs´alpha beträgt für diese Subskala .76 und liegt damit in dem von Harfst et al. (2004b) ermittelten Bereich (α = .77) (siehe Tabelle 18).

---

[12] Das Kriterium von mindestens 75% gültiger Werte zur Skalenbildung wurde generell für alle berechneten Skalen in der vorliegenden Arbeit zugrunde gelegt und folgt damit der Empfehlung von Wirtz (2004) zur Ersetzung fehlender Werte mittels des EM-Algorithmus, in der mindestens 70% gültiger Werte als Minimalanforderung definiert werden.

Weiterhin ging als Prädiktor der **„Body Mass Index“** ein **(BMI)**, der nach der Berechnungsvorschrift im ICD-10 (Dilling et al., 1993) aus den beiden Merkmalen „Größe“ und „Gewicht“ ermittelt wurde. Die Prozentzahl fehlender Werte für „Größe“ und „Gewicht“ beträgt 5.9%. Erwartungsgemäß ist die Variable „BMI“ deutlich rechtsschief verteilt und weist einen erhöhten Wert in der Kurtosis auf (siehe Tabelle 18).

Als weitere intervallskalierte Variable ging das Merkmal **„Alter“** in nachfolgende Analysen ein. Auch hier ergab eine deskriptive Analyse fehlender Werte einen eher geringen Anteil von 4.1% (siehe Tabelle 18).

**Zusammenfassend** kann festgehalten werden, dass die Überprüfung der Skalenstrukturen und der psychometrischen Eigenschaften insgesamt zufrieden stellende Ergebnisse erbracht haben, weswegen die Skalen der eingesetzten Selbsteinschätzungsinstrumente für nachfolgende Analysen verwendet wurden.

Nachfolgend soll eine Darstellung der Verteilungseigenschaften der standardisierten nominal- bzw. ordinalskalierten Items erfolgen, die als Prädiktoren in die Analysen zur Fallgruppenbildung aufgenommen werden sollen.

Tabelle 18: Psychometrische Kennwerte und Verteilungseigenschaften der eingesetzten Selbsteinschätzungsinstrumente und Items auf metrischem Niveau

| Items/Skalen der Instrumente | n | Fehlende Werte (%) | M | SD | Schiefe | SE der Schiefe | Kurtosis | SE der Kurtosis | Cronbach´s Alpha |
|---|---|---|---|---|---|---|---|---|---|
| Skalen der Selbsteinschätzungsinstrumente | | | | | | | | | |
| SF-8-S Psych. | 1888 | 1.6 | 2.26 | 0.89 | -0.31 | 0.06 | -0.57 | 0.11 | .80 |
| SF-8-S Som. | 1899 | 1.0 | 1.83 | 0.95 | 0.15 | 0.06 | -0.92 | 0.11 | .84 |
| SCL-14 DEP | 1876 | 2.2 | 1.82 | 1.05 | 0.16 | 0.06 | -0.92 | 0.11 | .89 |
| SCL-14 PHOB | 1869 | 2.6 | 0.54 | 0.79 | 1.78 | 0.06 | 2.81 | 0.11 | .82 |
| SCL-14 Som. | 1871 | 2.5 | 1.19 | 0.99 | 0.80 | 0.06 | -0.02 | 0.11 | .82 |
| SCL-14 GSI | 1866 | 2.7 | 1.27 | 0.74 | 0.61 | 0.06 | 0.10 | 0.11 | .88 |
| ADS-K Ges. | 1816 | 5.3 | 22.11 | 9.67 | -0.03 | 0.06 | -0.73 | 0.12 | .83 |
| IIP Subskala: „kühl und abweisend“ | 1819 | 5.2 | 1.08 | 0.80 | 0.75 | 0.06 | 0.27 | 0.12 | .76 |
| BMI | 1804 | 5.9 | 25.78 | 7.91 | 3.96 | 0.06 | 42.11 | 0.12 | |
| Alter | 1840 | 4.1 | 42.31 | 13.96 | -.076 | .057 | -0.61 | 0.11 | |

**Anmerkungen:** SF-8-S psych. = Psychische Summenskala der SF-8-Selbsteinschätzung; SF-8-S som. = Somatische Summenskala der SF-8-Selbsteinschätzung; SCL-14 DEP = Depressivitätsskala der SCL-14; SCL-14 PHOB = Phobische Angst der SCL-14; SCL-14 Som. = Somatisierungsskala der SCL-14; SCL-14 GSI = Global Severity Index der SCL-14; ADS-K Ges. = Gesamtwert der ADS-K; BMI = Body Mass Index

#### *9.5.1.2.6 Verteilungseigenschaften der selbst eingeschätzten Items mit nicht-metrischem Skalenniveau*

Die Abbildung 18 bis Abbildung 34 veranschaulichen die Verteilung der nominalskalierten Prädiktoren. Eine detaillierte Auflistung der Zellhäufigkeiten der einzelnen Items befindet sich im Anhang (vgl. Anhang 16.3.1). Dabei wurden die Merkmale hinsichtlich ihrer Variation gruppiert.

Es zeigt sich, dass 10 der 17 Prädiktoren als gut zu bewertende Verteilungseigenschaften bzw. Zellbesetzungen der einzelnen Kategorien aufweisen. Es handelt sich hierbei um die Variablen „Partnersituation“, „höchster beruflicher Abschluss“, „höchster Schulabschluss“, „aktuelle berufliche Situation“, „gegenwärtig berufliche Position“, „Arbeitsunfähigkeit/Krankschreibung zu Beginn der Rehabilitationsmaßnahme“, „Einnahme von Schmerzmitteln“, „Inanspruchnahme ambulanter Psychotherapie“, „Nikotinkonsum“ und „Arbeitsunfähigkeitszeiten für Berufstätige“ (siehe Abbildung 18 bis Abbildung 27).

Die deskriptive Analyse ergab weiterhin, dass - wie für Patienten in stationärer psychotherapeutischer Rehabilitationsbehandlung zu erwarten war - die Prädiktoren „Geschlecht“ und „Einnahme von Antidepressiva“ schief verteilt sind. Dabei nahmen 30% der Patienten täglich Medikamente ein, 40% hingegen keine (vergleiche Abbildung 28 und Abbildung 29).

Für weitere sechs Prädiktoren zeigte die deskriptive Analyse erwartungsgemäß eine unzureichende Variation in den Kategorien. Es handelt sich um die Variablen „Rentenabsicht“, „Antrag auf Erwerbsminderungsrente“, „Einnahme von Schlafmitteln“, „stationäre Vorbehandlungen“ und „Arbeitsunfähigkeitszeiten für Nicht-Berufstätige“. So trugen sich von den insgesamt 1918 Patienten 4% mit der Absicht, eine Rente zu beantragen und 5% der Patienten wollten einen Antrag auf Erwerbsminderungsrente stellen. 13% der Patienten hatten bereits vor der aktuellen stationären psychotherapeutischen Rehabilitationsmaßnahme einen psychotherapeutischen Krankenhausaufenthalt in der Krankenvorgeschichte (siehe Abbildung 30 bis Abbildung 34).

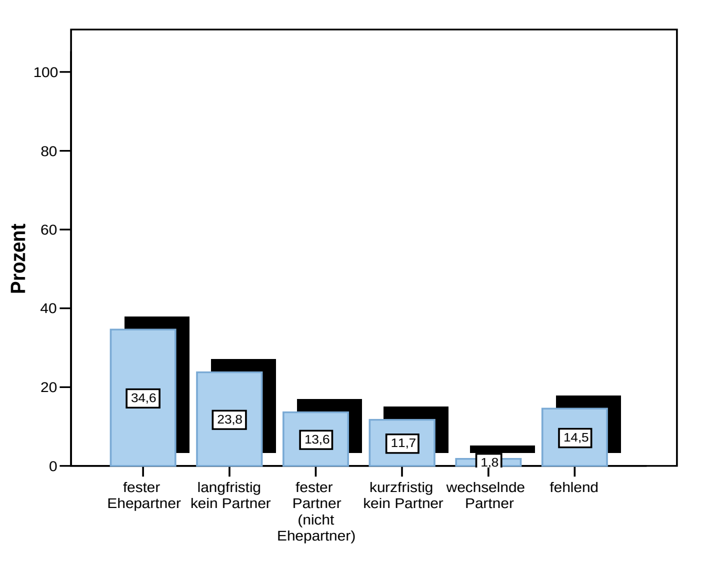



Abbildung 18: Verteilung des Prädiktors „Partnersituation“

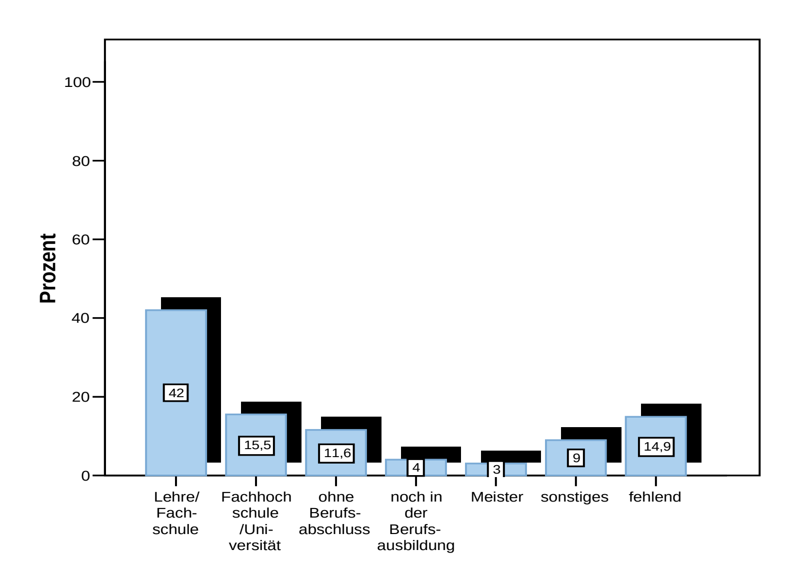



Abbildung 19: Verteilung des Prädiktors „höchster beruflicher Abschluss“

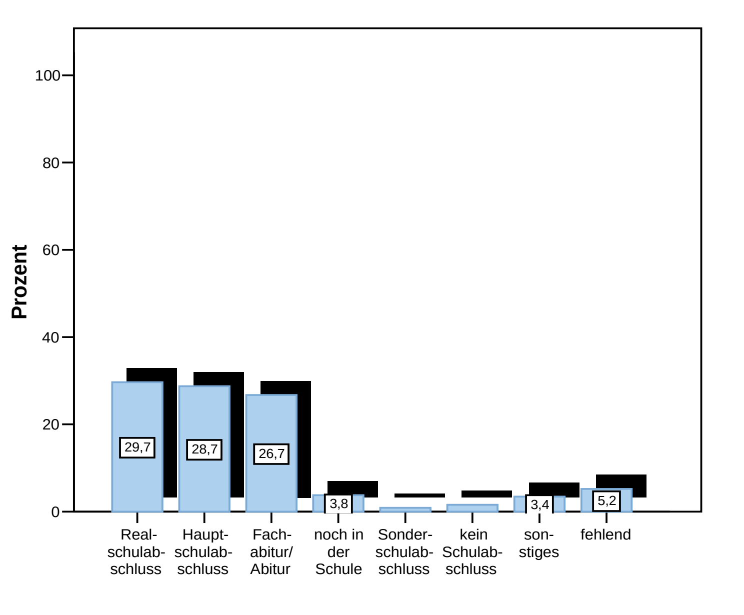



Abbildung 20: Verteilung des Prädiktors „höchster Schulabschluss“

Prozent
100
80
60
40
20
0
28,3
14,3
12,1
7,9
5,7
5,3
5,1
4,4
2
9,6
Vollzeit
Teilzeit
arbeitslos
nicht berufstätig
Berufsunfähigkeitsrente
Ausbildung
Pension
anderweitig
gelegentlich
unklar
nicht berufstätig
Zivildienst
beschützt
Frührente
Witwer
fehlend

Abbildung 21: Verteilung des Prädiktors „aktuelle berufliche Situation“

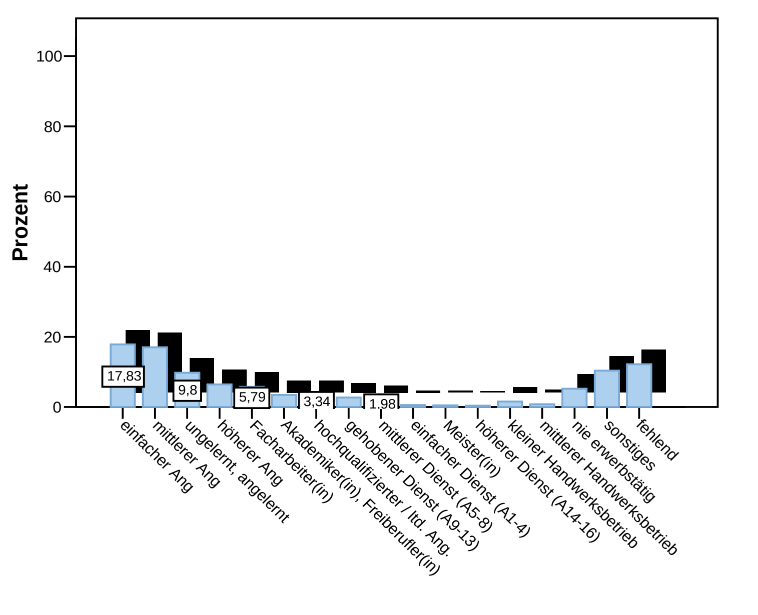



Abbildung 22: Verteilung des Prädiktors „gegenwärtige berufliche Position“

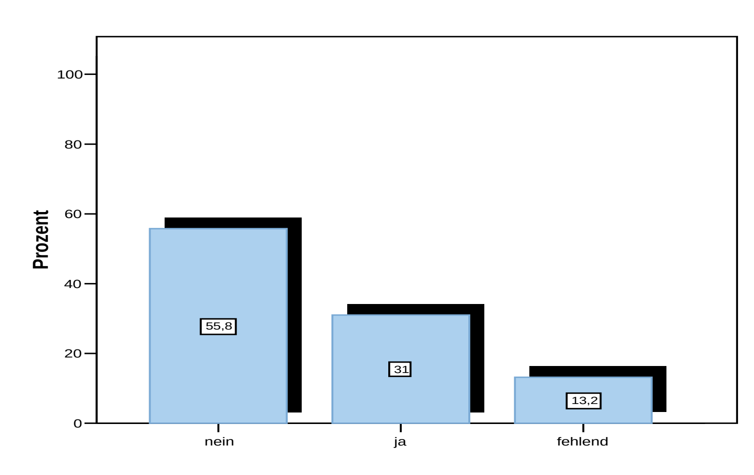



Abbildung 23: Verteilung des Prädiktors „Krankschreibung zu Beginn der Behandlung“

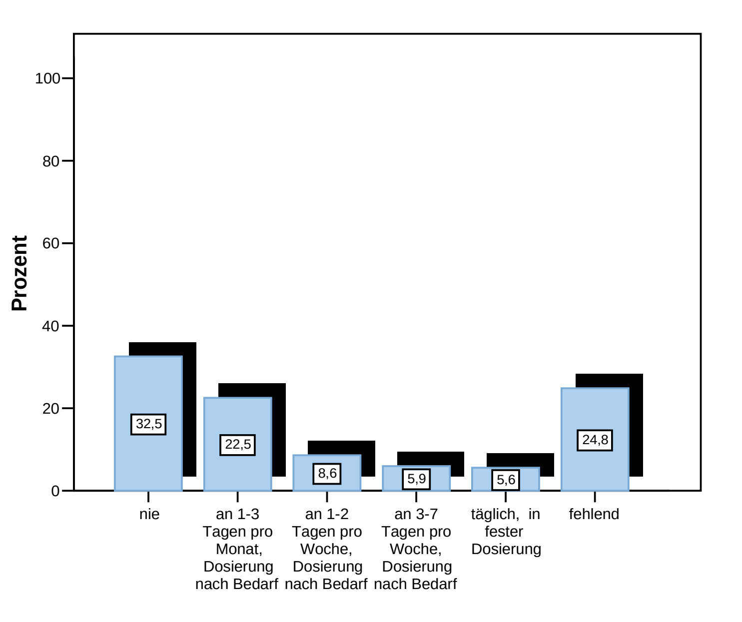



Abbildung 24: Verteilung des Prädiktors „Einnahme von Schmerzmitteln“

Prozent
100
80
60
40
20
0
35,1
24,2
15
11,9
13,8
nein
6-25 Termine
1-5 Termine
mehr als 25 Termine
fehlend

Abbildung 25: Verteilung des Prädiktors „Inanspruchnahme ambulanter Psychotherapie vor Aufnahme“

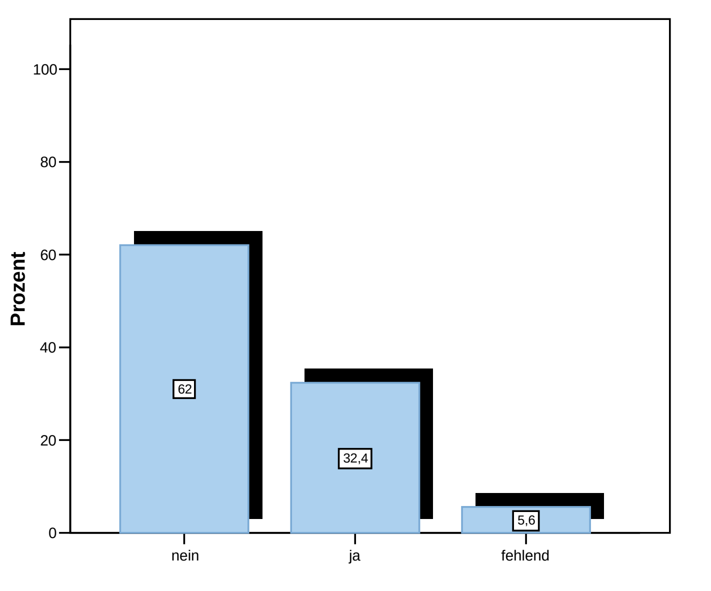



Abbildung 26: Verteilung des Prädiktors „Nikotinkonsum“

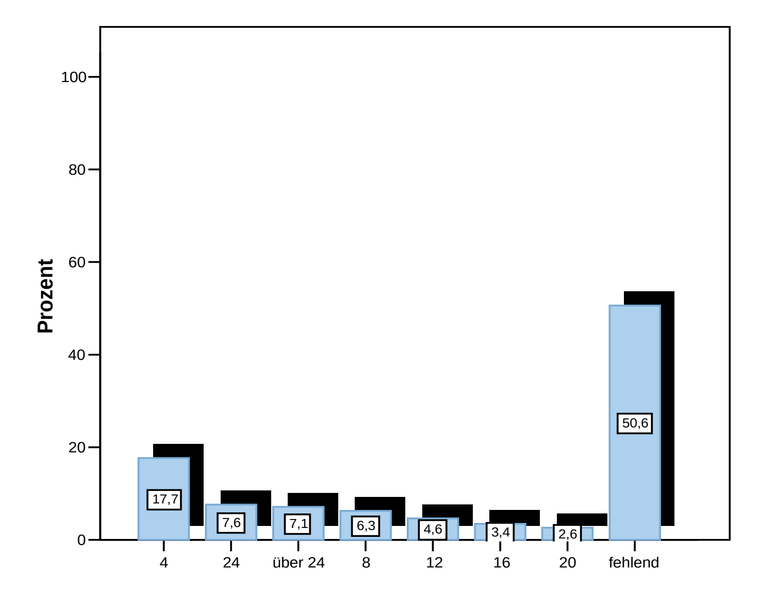



Abbildung 27: Verteilung des Prädiktors „Arbeitsunfähigkeitszeiten für Berufstätige“

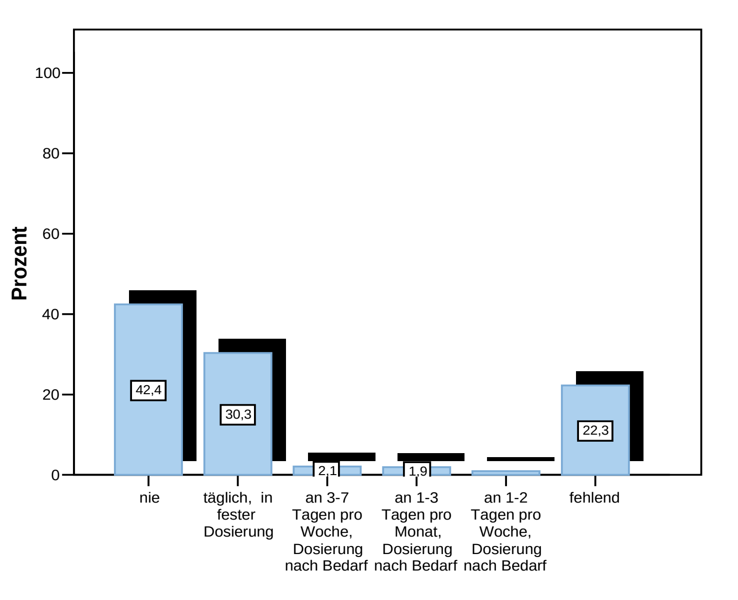



Abbildung 28: Verteilung des Prädiktors „Einnahme von Antidepressiva“

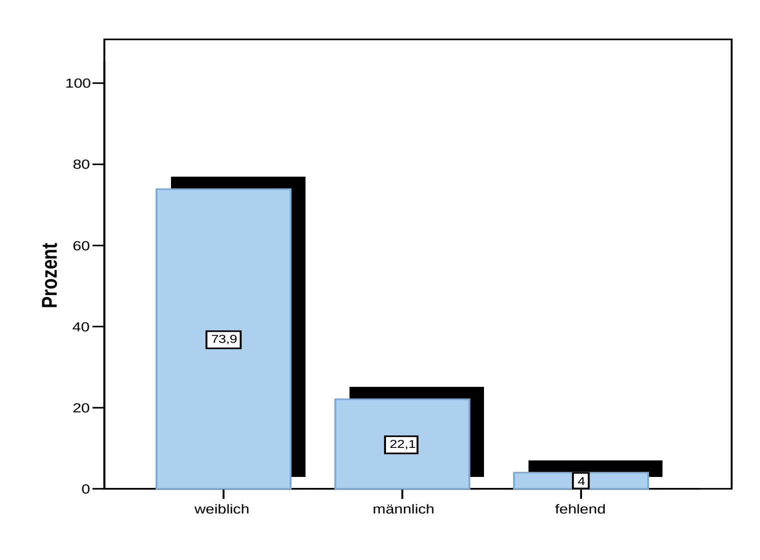



Abbildung 29: Verteilung des Prädiktors „Geschlecht“

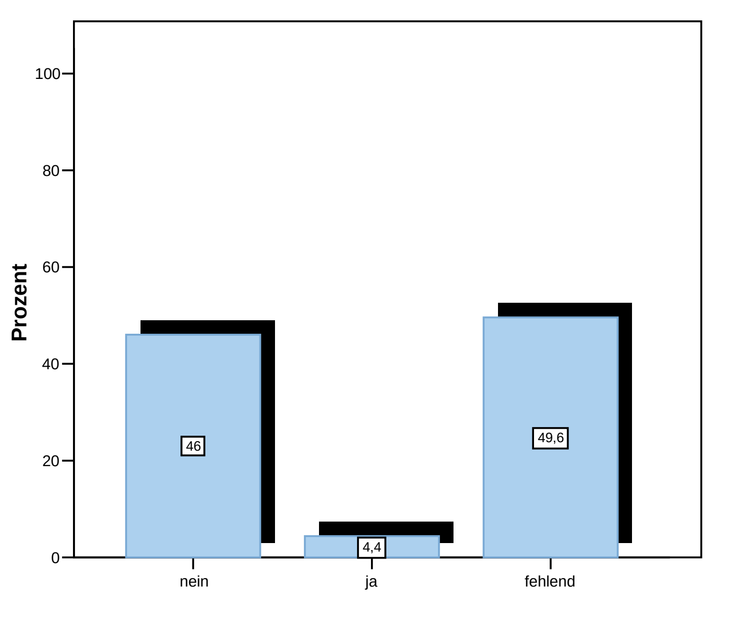



Abbildung 30: Verteilung des Prädiktors „Absicht einen Rentenantrag zu stellen“

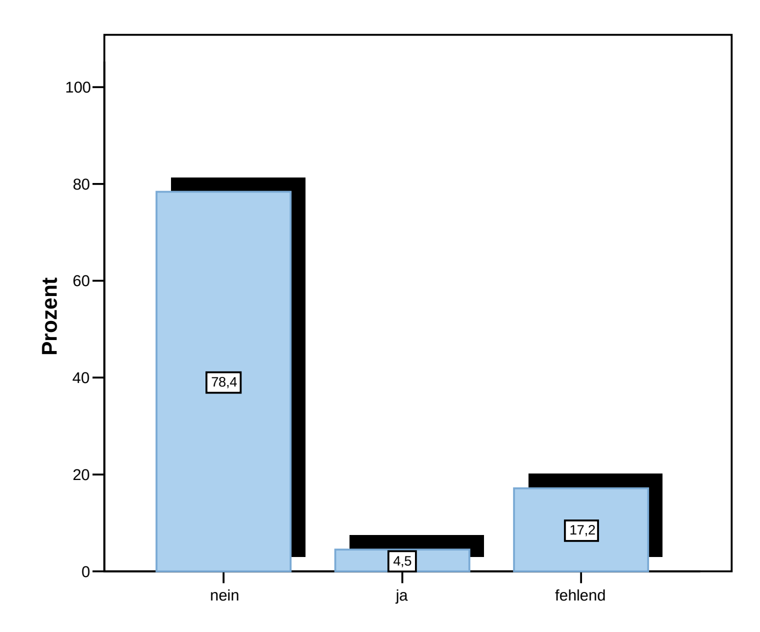



Abbildung 31: Verteilung des Prädiktors „Antrag auf Erwerbsminderungs-rente“

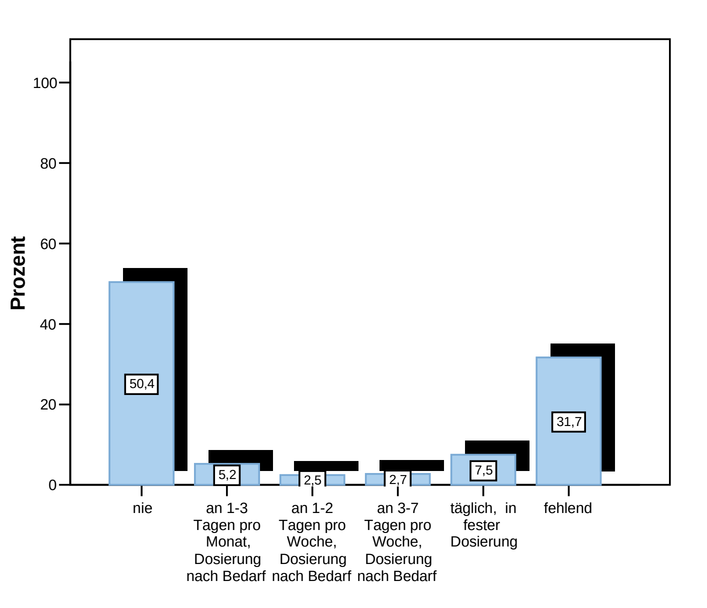



Abbildung 32: Verteilung des Prädiktors „Einnahme von Schlafmitteln“

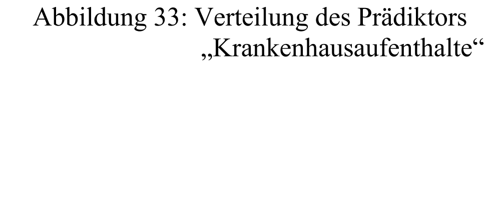



Abbildung 33: Verteilung des Prädiktors „Krankenhausaufenthalte“

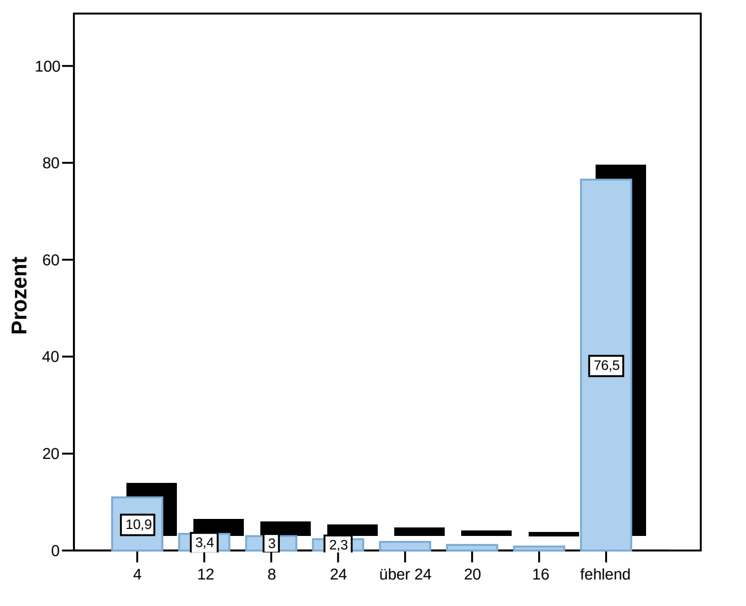



Abbildung 34: Verteilung des Prädiktors „Arbeitsunfähigkeitszeiten für Nicht-Berufstätige“

#### *9.5.2 Fremdeinschätzung zu Beginn der Behandlung*

In Anlehnung an die Darstellung der Selbsteinschätzungsinstrumente sollen im Folgenden die in der vorliegenden Untersuchung verwendeten Fremdeinschätzungsinstrumente und standardisierten Items beschrieben werden. Auch hier wurden zur Prüfung der Skalenstruktur konfirmatorische Faktorenanalysen und deskriptive Itemanalysen durchgeführt.

##### 9.5.2.1 Verwendete Fremdeinschätzungsinstrumente und Items

Eine detaillierte Auflistung der fremd einzuschätzenden Prädiktoren und ihrer Operationalisierungen findet sich in Tabelle 19. Wie bereits in Abschnitt 4.3 dargelegt und der Logik der allgemeinen DRG-Systeme folgend wurde die Diagnose von Patienten mit psychischen Störungen als ein Merkmal der krankheitsbezogenen Dimension mit aufgenommen. Die Operationalisierung wurde über die International Classification of Diseases (ICD-10, WHO, 1992) vorgenommen. Die Diagnosestellung erfolgte in der ersten Woche des stationären Aufenthaltes durch die Behandler der Patienten, die auch die Bezugstherapeuten waren.

Ein weiteres zentrales Kriterium der krankheitsbezogenen Dimension, die **Symptombelastung** der Patienten, soll über verschiedene standardisierte Instrumente und Items operationalisiert werden. Dazu zählt die deutsche Version der HoNOS zur differenzierten Erfassung des Schweregrades, deren Einsatz bereits unter Abschnitt 5.2 begründet wurde. Darüber hinaus soll die SF-8 in der vorliegenden Arbeit auch als Fremdeinschätzungsinstrument eingesetzt werden (Harfst et al. in Vorbereitung), wobei hinsichtlich Item und Antwortformat keine Unterschiede zur durch die Patienten selbst eingeschätzten Version bestehen. Neben diesen beiden standardisierten Instrumenten werden zwei weitere Items zur Erfassung des Symptomausmaßes verwendet. Zum einen handelt es sich dabei um ein standardisiertes Item, bei dem die Behandler die Erstdiagnose der Patienten hinsichtlich ihres Schweregrades auf einer Skala von 0 (= unauffällig) bis 3 (= schwer) einschätzen sollten. Zum anderen soll das Merkmal „Chronifizierung der Erkrankung“ in die Analysen einbezogen werden (siehe auch Anhang 16.2.1).

Weitere als ressourcenverbrauchsrelevant identifizierte Kriterien der krankheitsbezogenen Dimension (vgl. Abschnitt 4.1.2) sind die „Motivation zur Rehabilitationsbehandlung“ und sozialmedizinische Merkmale. Auch hier wurden beide Kriterien über standardisierte Items operationalisiert.

Tabelle 19: Prädiktoren der krankheitsbezogenen Dimension und deren Operationalisierungen (Fremdeinschätzung)

| **Dimension** | **Prädiktoren** | **Operationalisierung** |
|---|---|---|
| **Krankheitsbezogene Merkmale** | Diagnosen | • Klassifikationskategorien der ICD-10 (WHO, 1992)<br>• Komorbidität |
| | Symptombelastung | • Schweregradzusatz der Erstdiagnose<br>• Psychische und Somatische Summenskala der SF-8 (Harfst et al. in Vorbereitung)<br>• Gesamtscore der Health of the Nation Outcome Scales (HoNOS-D)<br>• Chronifizierung der Erkrankung |
| | Motivation zur Rehabilitationsbehandlung | • keine Motivation bis sehr hohe Motivation |
| | Sozialmedizinische Merkmale | • Art der Maßnahme<br>• Rentenantrag<br>• Rentner<br>• Arbeitsfähigkeit bei Aufnahme |

#### 9.5.2.2 Skalenüberprüfung der Fremdeinschätzungsinstrumente

Zur Prüfung der Skalenstruktur und faktoriellen Validität der eingesetzten Instrumente wurden zunächst konfirmatorische Faktorenanalysen berechnet, deren Ergebnisse im Folgenden dargestellt werden. Anschließend werden weitere psychometrische Kennwerte der Skalen berichtet. Die Überprüfung der psychometrischen Eigenschaften der Items und des Gesamtscores der HoNOS-D soll im Ergebnisteil vorliegender Arbeit dargestellt werden.

##### *9.5.2.2.1 Fragebogen zur gesundheitsbezogenen Lebensqualität-Fremdeinschätzung (SF-8-F)*

Analog zur Selbsteinschätzungsversion wurde in der vorliegenden Untersuchung die SF-8 als Fremdeinschätzungsverfahren (SF-8-F) verwendet (Harfst et al. in Vorbereitung). Da bislang noch keine psychometrischen Kennwerte für die SF-8-F vorliegen, sollen im Folgenden die Ergebnisse der konfirmatorischen Faktorenanalyse berichtet und nachfolgend psychometrische Kennwerte der Skalen vorgestellt werden. Die Modellspezifikation orientierte sich hierbei an der Skalenbildung der Selbsteinschätzungsversion der SF-8 (SF-8-S).

Abbildung 35 gibt einen Überblick über das postulierte Modell. Die standardisierten Pfadkoeffizienten weisen durchwegs Werte über .50 auf und können damit als zufrieden stellend bewertet werden. Die Ergebnisse der gewählten Anpassungstests der KFA bestätigen nur teilweise die theoretisch angenommene Skalenstruktur. Der SRMR ist zwar kleiner als .09 (SRMR = .0854), der RMSEA liegt aber außerhalb des Cut-off Bereiches von .06 und außerhalb des gerade noch akzeptablen Bereiches mit .11 (RMSEA = .12).

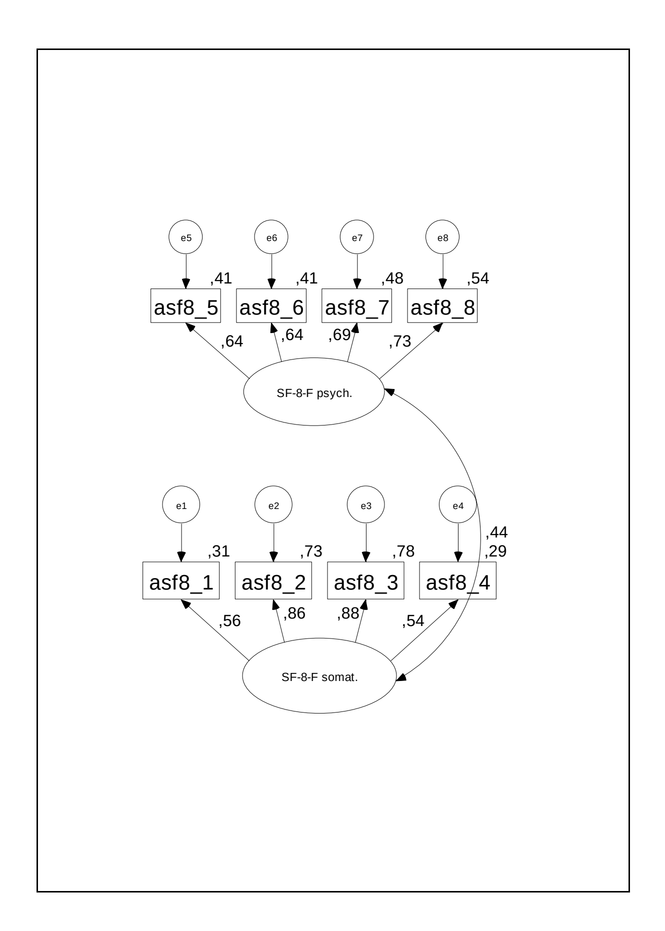



Abbildung 35: Postulierte und mittels konfirmatorischer Faktorenanalyse überprüfte Skalenstruktur der SF-8-Fremdeinschätzung

#### 9.5.2.2.2 *Psychometrische Kennwerte der verwendeten Fremdeinschätzungsinstrumente und Items mit metrischen Skalenniveau*

Wie bereits in Abschnitt 9.5.1.2.2 dargelegt, erfolgte die Skalenbildung der SF-8-Fremdeinschätzung in Analogie zur SF-8-Selbsteinschätzung.

Eine deskriptive Analyse fehlender Werte für die Skalen der SF-8-F bzw. die standardisierten Items zeigt Ausfallquoten in einem zufrieden stellenden Bereich zwischen 0.8% und 2.7% (vergleiche Tabelle 20). Auch die Schiefe- und Kurtosisindizes liegen in einem akzeptablen Bereich unter 1 und sprechen für eine annähernde Normalverteilung der eingesetzten Skalen. Cronbach´s alpha nimmt für die Subskalen der SF-8 befriedigende Werte von .77 und .78 ein.

Tabelle 20: Psychometrische Kennwerte und Verteilungseigenschaften der eingesetzten Fremdeinschätzungsinstrumente und standardisierten Items auf metrischem Skalenniveau

| **Items/Skalen der Instrumente** | **n** | **Prozentzahl fehlender Werte** | **M** | **SD** | **Schiefe** | **SE der Schiefe** | **Kurtosis** | **SE der Kurtosis** | **Cronbach´s Alpha** |
|---|---|---|---|---|---|---|---|---|---|
| Fremdeinschätzung | | | | | | | | | |
| SF-8-F Psychische Summenskala | 1901 | 0.9 | 2.59 | 0.65 | -0.83 | 0.06 | 0.82 | 0.12 | .77 |
| SF-8-F Somatische Summenskala | 1902 | 0.8 | 1.80 | 0.81 | 0.05 | 0.06 | -0.84 | 0.11 | .78 |
| Chronifizierung in Jahren | 1867 | 2.7 | 3.4 | 1.44 | 0.02 | 0.06 | -0.94 | 0.11 | |
| Motivation zur Rehabilitationsbehandlung | 1869 | 2.6 | 2.58 | 1.16 | 0.68 | 0.06 | -0.05 | 0.11 | |

#### 9.5.2.2.3 *Verteilungseigenschaften der fremd eingeschätzten Prädiktoren mit nicht-metrischem Skalenniveau*

Die Abbildung 36 bis Abbildung 39 veranschaulichen die Verteilung der nominalskalierten Variablen. Eine detaillierte Auflistung der Zellbesetzungen in den einzelnen Kategorien der Kriterien befindet sich im Anhang 16.3.2. Hierbei weisen die Variablen „Arbeitsfähigkeit bei Aufnahme“ und „Schweregradeinschätzung der Erstdiagnose“ eine gute Zellbesetzung der einzelnen Kategorien auf. Erwartungsgemäß und in Analogie zu den selbst eingeschätzten Items zeigt sich auch hier, dass die Prädiktoren „Rentenantrag“ und „Rentner“ schiefe Verteilungseigenschaften aufweisen: 5% der Patienten hatten einen Rentenantrag gestellt und 14% der Patienten waren Rentner.

Auch für den Prädiktor „Chronifizierung der Erkrankung“ ergeben sich aus der Literatur bekannte Verteilungseigenschaften. Am häufigsten litten die Patienten bereits seit 3 Jahren unter ihrer psychischen Erkrankung, mit einer Spannweite von 2 bis 5 Jahren (vgl. Abbildung 40).

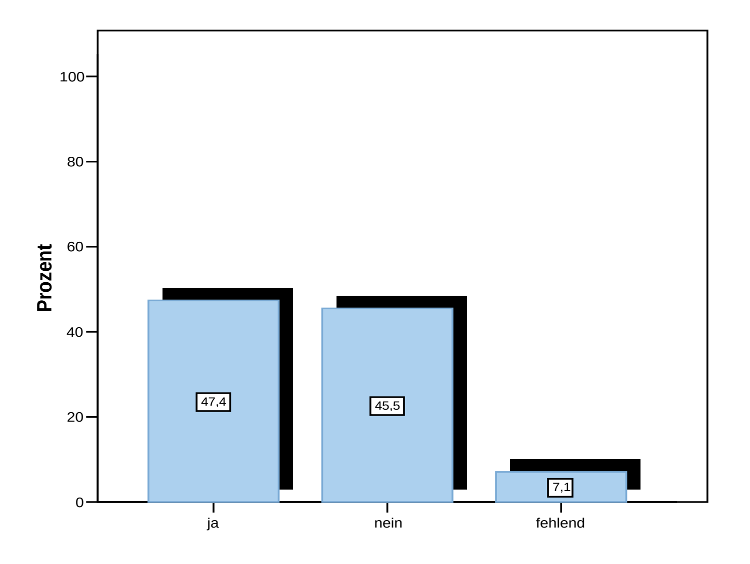



Abbildung 36: Verteilung des Prädiktors „Arbeitsfähigkeit bei Aufnahme“

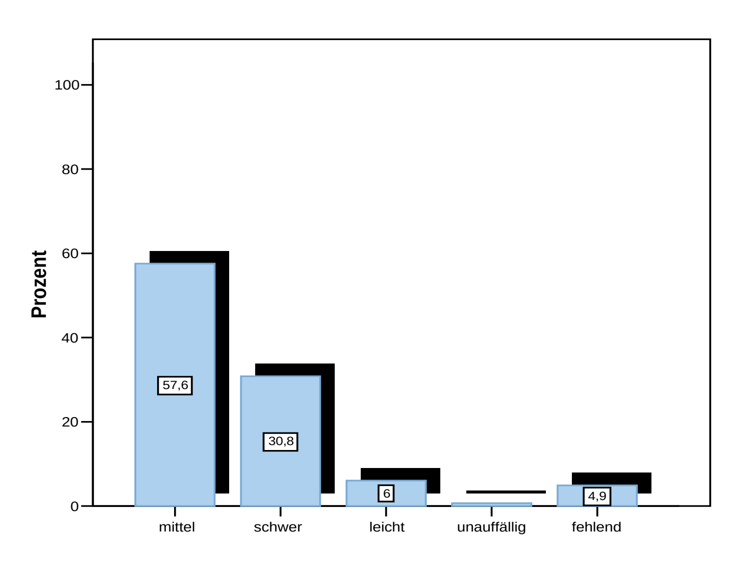



Abbildung 37: Verteilung des Prädiktors „Schweregradzusatz der Erstdiagnose“

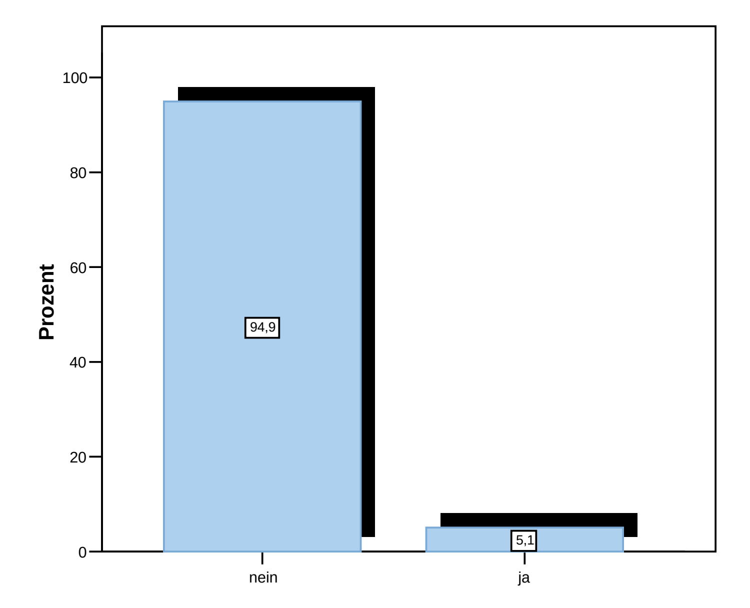



Abbildung 38: Verteilung des Prädiktors „Rentenantrag“

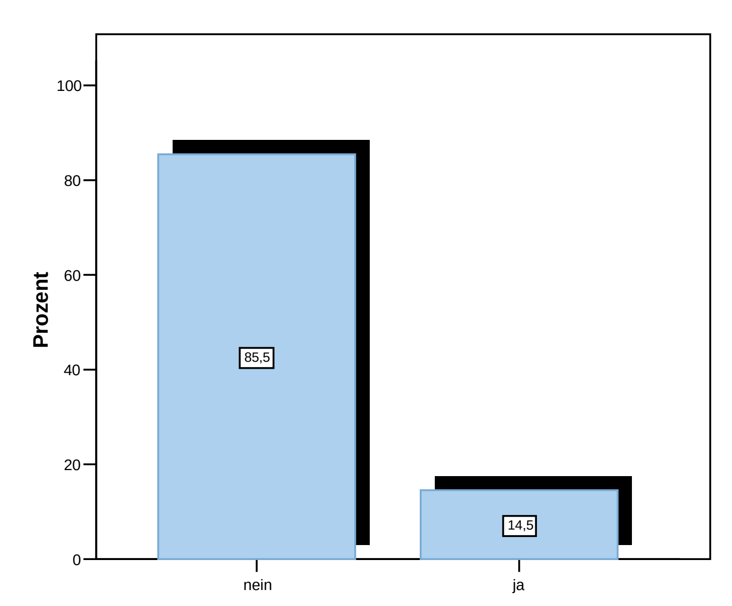



Abbildung 39: Verteilung des Prädiktors „Rentner“

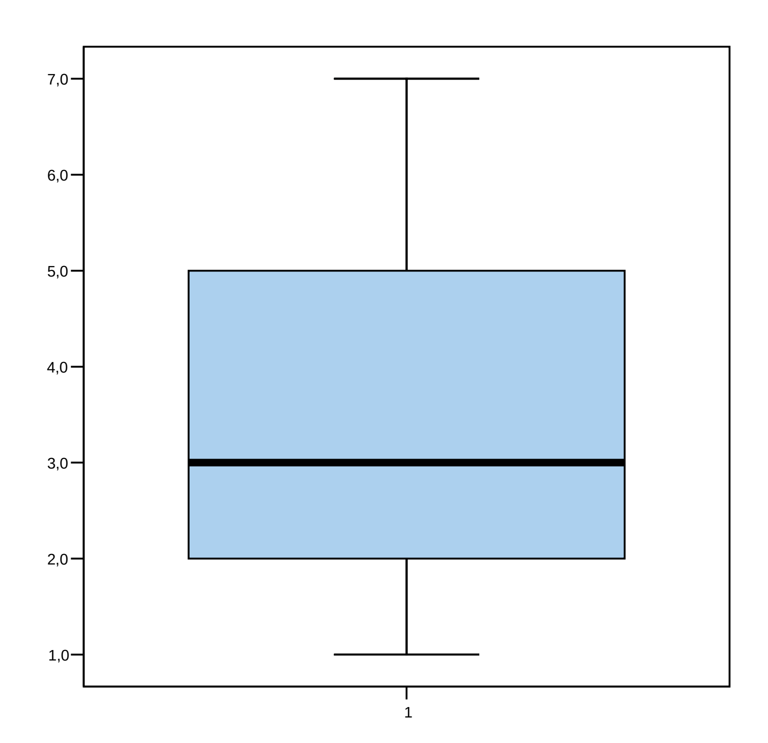



Abbildung 40: Verteilung des Prädiktors „Chronifizierungsdauer der Haupterkrankung“ (Ordinalskalenniveau)

**Zusammenfassend** lässt sich festhalten, dass die fremd eingeschätzten Prädiktoren trotz der größtenteils schiefen Verteilungseigenschaften zum einen ein Mindestkriterium an Variation erfüllen sowie theoretisch und empirisch von großer Bedeutung sind, weswegen sie in die nachfolgenden Analysen aufgenommen werden.

Nachdem die psychometrischen Kennwerte der Prädiktoren dargestellt wurden, sollen im Folgenden die Verteilungseigenschaften der **Kriteriumsvariable** bzw. **abhängigen Variable „Behandlungsdauer“** vorgestellt werden.

### *9.5.3 Die Erfassung der Kriteriumsvariable „ Behandlungsdauer“*

Die Behandlungsdauer nimmt innerhalb der Fallgruppenbildung von Patienten eine zentrale Stellung zur Vorhersage des Ressourcenverbrauchs ein. So erfolgt die Klassifikation von Patienten in allgemeinen DRG-Systemen (z.B. AR-DRG-System) in Abhängigkeit von deren Aufenthaltsdauer. Auch in den komplementär zu den allgemeinen DRG-Systemen entwickel-

ten internationalen Klassifikationssystemen wird in der überwiegenden Anzahl (6 von 9) die Behandlungs- bzw. Verweildauer zur Vorhersage des Ressourcenverbrauchs zugrunde gelegt (vergleiche Tabelle 2 in Abschnitt „3.2“). In Abgrenzung zur *Verweildauer*, welche sich nach der Bundespflegesatzverordnung durch die Anzahl der Tage ergibt, die ein Patient stationär im Krankenhaus behandelt wird (einschließlich Aufnahme- und Entlassungstag), soll in der vorliegenden Arbeit die **Behandlungsdauer** als ein zentrales Bestimmungsstück des Ressourcenverbrauchs verwendet werden. Dabei fußt die Behandlungsdauer zwar auf der gleichen Berechnungsgrundlage wie die Verweildauer, allerdings mit der Ausnahme, dass Patienten die zur Probe nach Hause entlassen und anschließend wieder aufgenommen werden, nicht als neuer Fall gezählt, sondern die Behandlungstage des ersten und zweiten Aufenthaltes aufaddiert werden (Tritt et al., 2002).

In Bezug auf das einzusetzende Regression-Tree-Verfahren sollte die abhängige Variable „Behandlungsdauer“ einige methodische Voraussetzungen erfüllen. Dementsprechend sollten zunächst die Verteilungseigenschaften der Kriteriumsvariablen in der Stichprobe überprüft werden.

Abbildung 41 zeigt die **Verteilung der Behandlungsdauer (BD)** in der Untersuchungsstichprobe. Die Kennwerte der BD sind in Tabelle 21 zusammen getragen. Von 11 Patienten der Untersuchungsstichprobe (N = 1918) lagen keine Angaben zur Behandlungsdauer vor, weswegen sie von den weiteren Analysen ausgeschlossen worden (n = 1907). Aufgrund der geringen Anzahl von Patienten ohne gültige Werte in der BD (n = 10) wurde auf Überprüfung systematischer Unterschiede hinsichtlich klinischer und soziodemographischer Merkmale in Bezug zur Gruppe der Patienten mit gültigen Werten in der BD verzichtet.

Die durchschnittliche Gesamtbehandlungsdauer betrug über alle Patienten 47,5 Tage (SD = 17,4 Tage). Der Median lag bei 42 Tagen. An dieser Stelle kann bereits angemerkt werden, dass damit die Dauer der Behandlung in den 11 Kliniken über der durchschnittlichen Verweildauer von 38,2 Tagen (Schulz et al., in Druck) für Patienten in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung liegt.

Darüber hinaus wurde zur weiteren Charakterisierung der Verteilung der BD der **Variationskoeffizient** über alle 11 Kliniken berechnet. Dieser ergibt sich aus der Relation von Standardabweichung und arithmetischem Mittel (Bortz, 1995) und beträgt für die BD in der vorliegen-

den Stichprobe 36,7%. In Anlehnung an Fischer (2003) können Variationskoeffizienten über 50% als relativ hoch interpretiert werden, d.h. als Heterogenität anzeigende Werte; Variationskoeffizienten unter 30% weisen dagegen auf homogenere Verteilungen hin. Dementsprechend ist die Verteilung der Kriteriumsvariablen „BD" als moderat heterogen einzustufen.

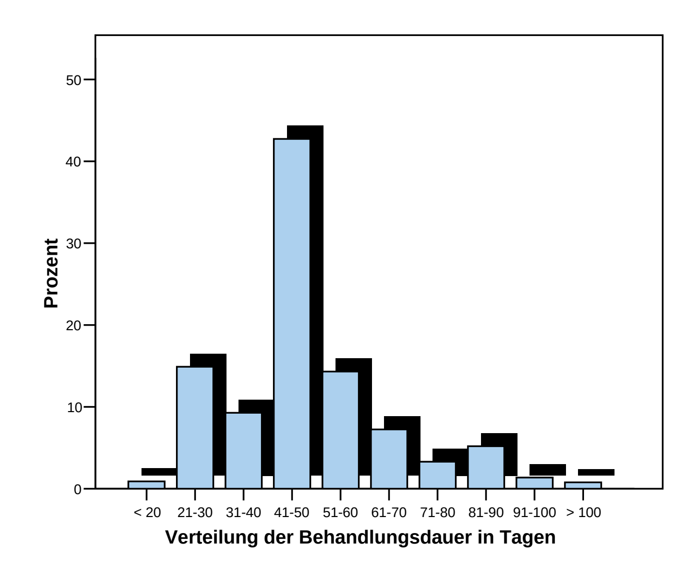



Abbildung 41: Häufigkeitsverteilung der fallbezogenen Behandlungsdauer in Tagen (Kategorien)

Tabelle 21: Kennwerte der Kriteriumsvariablen „Behandlungsdauer" für die Untersuchungsstichprobe (n = 1907)

| | **M** | **SD** | **Mdn** | **Schiefe** | **SE der Schiefe** | **Kurtosis** | **SE der Kurtosis** |
|---|---|---|---|---|---|---|---|
| Behandlungsdauer (n = 1907) | 47.5 | 17.42 | 42 | 1.00 | .056 | 1.07 | .112 |

Weiterhin sollte an dieser Stelle noch darauf hingewiesen werden, dass in der Literatur unter methodischen und mathematischen Gesichtspunkten häufig zwei Vorgehensweisen („Trimmen" und „Logarithmieren") zum Umgang mit der BD aufgeführt werden (vergleiche hierzu auch Richter, 2001). Im erstgenannten Fall, dem sog. „Trimmen", werden Ausreißer entweder aufgrund eines festgesetzten Wertes (z.B. Liegedauer > 100 Tage) oder aufgrund statistischer

Methoden (z.B. Eliminierung von Extremwerten mittels explorativer Datenanalyse) von der weiteren Berechnung ausgeschlossen. Dieses Vorgehen hat den Vorteil, zur besseren Interpretierbarkeit der Ergebnisse beizutragen, allerdings bringt es unter einem klinisch-anwendungsbezogenen Gesichtspunkt auch den Nachteil des Informationsverlustes mit sich. In der vorliegenden Arbeit wurde zugunsten der klinisch-anwendungsbezogenen Perspektive von einem „Trimmen" der Kriteriumsvariable „BD" abgesehen, auch von der Annahme ausgehend, dass sich sog. „Ausreißer" (Patienten, die besonders lange Behandlungsdauern aufweisen) über die Kliniken zufällig verteilen. Ein weiteres methodisches Vorgehen bezieht sich auf die Empfehlung, die Werte der BD logarithmisch zu transformieren (Richter, 2001). Dies erlaubt u.a. unter der Prämisse der Normalverteilung der Kriteriumsvariablen einen besseren mathematischen Fit. Damit ist allerdings der Nachteil verbunden, dass die Ergebnisse nicht unmittelbar interpretierbar sind, weswegen im vorliegenden Fall zugunsten einer besseren Interpretierbarkeit entschieden und die abhängige Variable „BD" nicht logarithmisiert wurde.

## 9.6 Auswertungsmethodik

### *9.6.1 Methodische Verfahren zur Überprüfung der generellen Eignung der HoNOS-D (Hypothese 1a und 1b)*

Zur Überprüfung der *Hypothese 1a* „Praktikabilität" werden deskriptive Analysen (Ermittlung gültiger bzw. fehlender Werte, Berechnung von Mittelwerten, Standardabweichungen, prozentualen Häufigkeiten) der HoNOS-D Items in der Untersuchungsstichprobe vorgenommen.

Zur Überprüfung der *Hypothese 1b* „Verteilungseigenschaften" *und 1c* „Interkorrelation" sollen unterschiedliche statistische Auswertungsverfahren zur Anwendung kommen. Neben deskriptiven Analysen zu psychometrischen Kennwerten der Items der HoNOS-D (Mittelwerte, Standardabweichungen, Schiefe und Kurtosis), werden Interkorrelationen zwischen den Items berechnet. Aufgrund der erwarteten schiefen Verteilung einiger Items werden die Interkorrelationen mithilfe von Spearman-Rangkorrelationen berechnet.

Zur Überprüfung der *Hypothesen 1d* „Interraterreliabilität" *und 1e* „Test-Retest-Reliabilität" sollen Interraterreliabilitäten und Test-Retest-Reliabilitäten berechnet werden. Die Entschei-

dungen zur Wahl des geeigneten Koeffizienten für die Analysen sollen im Abschnit 9.6.1.1 näher erläutert werden.

Hinsichtlich des Signifikanzniveaus werden gemäß den konventionellen Richtlinien Irrtumswahrscheinlichkeiten von $\alpha \leq .05$ als signifikant bewertet. Zur Interpretation der Effektstärken werden die Konventionen von Cohen (1988) zugrunde gelegt: $d \geq .20$ (kleiner Effekt), $d \geq .50$ (mittlerer Effekt), $d \geq .80$ (großer Effekt), $r \geq .10$ (kleiner Effekt), $r \geq .30$ (mittlerer Effekt), $r \geq .50$ (großer Effekt).

#### 9.6.1.1 Methodisches Verfahren zur Überprüfung der Interrater-Reliabilitäten

In der Literatur wird in Abhängigkeit vom Skalenniveau eine Vielzahl an Koeffizienten zur Berechnung von Übereinstimmungen zwischen Beurteilungen verschiedener Personen diskutiert (Asendorpf & Wallbott, 1979; Wirtz & Caspar, 2002). Dabei geht die Wahl eines geeigneten Koeffizienten meist mit mehreren Entscheidungen einher.

Bei den Items der HoNOS-D handelt es sich um intervallskalierte Ratingskalen. Folgt man den Ausführungen von Wirtz und Caspar (2002), stellen Intraklassenkorrelationskoeffizienten (ICC) eine angemessene Methode zur Bestimmung von Interrater-Reliabilitäten bei vorliegendem Intervallskalenniveau dar, die im Übrigen auch in der Mehrzahl der Studien zu psychometrischen Eigenschaften der HoNOS Verwendung findet (vgl. hierzu Abschnitt 6.3). Dabei basieren Intraklassenkorrelationen auf einem varianzanalytischen Modell.

Eine nächste Entscheidung bezieht sich auf die Wahl des zugrunde liegenden varianzanalytischen Modells (1- vs. 2-faktorielles Modell). Das 1-faktorielle Modell zur Berechnung von Intraklassenkorrelationen ist dann zu verwenden, wenn die untersuchten Personen nicht alle von denselben Ratern beurteilt wurden. Im Gegensatz dazu wird das 2-faktorielle Modell verwendet, wenn alle Personen von denselben Ratern beurteilt wurden (Wirtz & Caspar, 2002). Da im vorliegenden Fall Beurteilungen von jeweils zwei Ratern miteinander verglichen werden und die Patienten nicht alle von denselben Ratern beurteilt wurden, soll das *1-faktorielle Modell* zur Berechnung der Interklassenkorrelationen verwendet werden. Damit werden alle Varianzanteile, die nicht auf den Unterschied zwischen den wahren Werten der Personen zurückzuführen sind als Teil der Fehlervarianz angesehen. Beim 1-faktoriellen Mo-

dell steht, im Gegensatz zum 2-faktoriellen Modell, nur das Modell der *unjustierten* Intraklassenkoeffizienten zur Verfügung. Im Vergleich zum adjustierten Modell handelt es sich hierbei um das strengere Modell, da auch Mittelwertsunterschiede zwischen den Ratern zulasten der Reliabilitätsschätzung gehen. So kann schließlich ein perfekter Zusammenhang (ICC = 1) nur dann erreicht werden, wenn zusätzlich zur Varianzhomogenität die Gleichheit der Mittelwerte gegeben ist. Definitionsgemäß sollte der ICC nur Werte zwischen 0 und 1 annehmen, in speziellen Fällen können allerdings auch negative Werte erzielt werden, die dann mit einer Reliabilität von 0 gewertet werden müssen (Wirtz & Caspar, 2002).

Eine letzte Entscheidung bei der Wahl des geeigneten Modells bezieht sich auf die Möglichkeit der Generalisierbarkeit der Ergebnisse. Sollen die untersuchten Rater eine repräsentative Stichprobe aller Rater darstellen, für die die Reliabilitätsaussage Gültigkeit besitzen soll, so sollte das Modell „random“ gewählt werden. Dagegen sollte das Modell Rater „fixed“ gewählt werden, wenn die Reliabilitätsschätzung lediglich eine Eigenschaft der tatsächlich untersuchten Raterstichprobe darstellt. Im vorliegenden Fall fiel die Entscheidung auf die Berechnung eines $ICC_{unjust}$ auf der Basis des 1-faktoriellen Modells mit Rater „random“. Bei der Beurteilung der Höhe der Intraklassenkorrelation können nach Wirtz und Caspar (2002) Werte ab $\geq 0.7$ als gut bezeichnet werden.

Problematisch bei der Wahl des ICC als Maß zur Beurteilung der Interrater-Reliabilität ist, dass dieser nur zuverlässige Werte liefert, wenn ausreichend Variation in den Mittelwerten der Objekte vorliegt (Asendorpf & Wallbott, 1979; Wirtz & Caspar, 2002). Da aber aufgrund der empirischen Befunde aus der Literatur schiefe Verteilungseigenschaften der Items erwartet werden müssen und damit vermutlich keine ausreichende Varianz in den Mittelwerten der Objekte vorliegt, wurden ergänzend alternative Strategien zur Berechnung des Ausmaßes der Beurteilerübereinstimmung herangezogen. Dabei handelt es sich zum einen um die **Bestimmung der prozentualen Übereinstimmung** und zum anderen um den **Finn-Koeffizienten**.

Die **prozentuale Übereinstimmung** gibt den prozentualen Anteil der Fälle an, bei denen beide Rater das gleiche Urteil abgegeben haben und ist wie folgt definiert (Wirtz & Caspar, 2002):

$$PÜ_{gesamt} = \frac{\text{Anzahl der von allen Ratern gleich beurteilten Personen}}{\text{Anzahl der insgesamt beurteilten Personen}} * 100\%$$

Zur Anwendung des **Finn-Koeffizienten** gehen in der Literatur die Meinungen weit auseinander. Während Asendorpf und Wallbott (1979), Tinsley und Weiss (1975) bei zu geringer Variation der Itemmittelwerte die Berechnung des Finn-Koeffizienten empfehlen, sprechen sich Wirtz und Caspar (2002) bei gegebener Voraussetzung des Intervallskalenniveaus in jedem Falle für den Einsatz des Intraklassenkoeffizienten aus. Bei Bewertung des Finn-Koeffizienten können Werte zwischen 0.50 – 0.70 als zufrieden stellend und Werte über .70 als gut bezeichnet werden.

Da jeder der oben erläuterten Koeffizienten Vor- und Nachteile mit sich bringt, wurde eine Kombination der Strategien gewählt. Es sollen in der vorliegenden Arbeit sowohl die ICC als auch die prozentuale Übereinstimmung und die Finn-Koeffizienten berechnet werden. Die Beurteilung der Güte der Interrater-Reliabilitäten soll wie folgt festgelegt werden: Sobald ein Wert der drei berichteten Koeffizienten innerhalb des akzeptablen Bereiches liegt, wird eine ausreichende Interrater-Reliabilität für das Item der HoNOS-D angenommen.

Der Finn-Koeffizient wurde nach folgender Formel berechnet (Wirtz & Caspar, 2002):

$$r_{Finn} = 1 - \frac{MS_{beobachtet}}{(k^2 - 1) / 12}$$

k = Anzahl der Kategorien

Da die Berechnung der ICC in einer Teilstichprobe pro Raterpaar erfolgt und damit verschiedene Stichprobengrößen zugrunde liegen, soll eine Vergleichbarkeit der verschiedenen Intraklassenkoeffizienten (ICC) durch eine Fisher-Z-Transformation hergestellt und für die Gesamtwerte vorgenommen werden. Durch dieses methodische Vorgehen werden die Korrelationskoeffizienten auf einer Verhältnisskala abgebildet (Wirtz & Caspar, 2002).

#### 9.6.1.2 Methodische Verfahren zur Überprüfung der Test-Retest-Reliabilität

In Anlehnung an das methodische Vorgehen der Testautoren Wing et al. (1998) und um eine internationale Vergleichbarkeit der Ergebnisse vorliegender Arbeit mit empirischen Befunden zum HoNOS zu gewährleisten, sollen zur Überprüfung der Hypothese 1e die Test-Retest-Reliabilitäten mittels Intraklassenkoeffizienten (Asendorpf & Wallbott, 1979) zwischen den Einschätzungen zu Beginn und am Ende der Therapie bestimmt werden. Da während dieser zwei Messzeitpunkte eine psychotherapeutische Behandlung durchgeführt wurde, soll nur die

Test-Retest-Reliabilität einer in sich zentralen Maßen des Therapieergebnisses nicht signifikant veränderten Subgruppe berechnet werden (n = 540). Als statistischer Kennwert einer signifikanten Veränderung durch die psychotherapeutische stationäre Behandlung wird der „Reliable Change Index“ (RCI, Jacobson & Truax, 1991) zugrunde gelegt. Als zentraler Outcomeparameter wurde der RCI aller Patienten der Stichprobe für die Subskala „Psychische Summenskala der SF-8-S und SF-8-F“ nach folgender Formel berechnet:

$RCI = 2 * \sqrt{2} * SD_1 * \sqrt{(1 - \mathrm{Rel})}$ (Jacobson & Truax, 1991)

(Anmerkung: $SD_1$ = Standardabweichung der jeweiligen zugrunde liegenden Untersuchungsstichprobe zu Messzeitpunkt $t_1$; Rel = interne Konsistenz der SF-8-Selbst- und Fremdeinschätzung in der zugrunde liegenden Untersuchungsstichprobe)

Daraus ergaben sich folgende Werte:

$$RCI_{SF-8-F} = 2 * \sqrt{2} * 0.648 * \sqrt{(1 - 0.769)} = 0.88$$

$$RCI_{SF-8-S} = 2 * \sqrt{2} * 0.892 * \sqrt{(1 - 0.799)} = 1.13$$

Die Subgruppe „Nicht-verändert“ wurde wie folgt definiert:

$RCI_{SF\text{-}8\text{-}F} \leq 0.88$ und $RCI_{SF\text{-}8\text{-}S} \leq 1.13$

#### *9.6.2 Methodische Verfahren zur Überprüfung der spezifischen Eignung der HoNOS-D*

Zur Überprüfung der Hypothese 2a „Prädiktor HoNOS-D“ wird ein Regression-Tree-Verfahren (Entscheidungsbaumverfahren) zur Vorhersage der Behandlungsdauer gerechnet. Wird die Hypothese 2a bestätigt, so soll zur Überprüfung der Hypothese 2b „Bedeutsamkeit des Prädiktors HoNOS-D“ eine Lineare Regressionsanalyse und zur Prüfung der Hypothese 2c „Position im Prädiktoren-Kern-Set“ ein Regression-Tree-Verfahren zur Anwendung kommen.

Da es sich beim Regression-Tree-Verfahren um eine im Rahmen der Beantwortung psychologisch-klinischer Fragestellungen relativ selten verwendete Methode handelt, soll sie im Folgenden etwas ausführlicher dargestellt werden.

#### 9.6.2.1 Das Regression-Tree-Verfahren

Zur Überprüfung der Hypothese 2a bis 2c soll ein **Regression-Tree-Verfahren** mittels des SPSS-Moduls „AnswerTree“ verwendet werden (SPSS, 2004). Dieses Segmentierungs- und Klassifikationsverfahren verfolgt die Zielsetzung, eine Population anhand einer Auswahl von Prädiktoren (Patientenmerkmale) in sich gegenseitig ausschließende, homogene Subgruppen zu untergliedern, die sich hinsichtlich eines vorgegebenen Kriteriums (Ressourceneinsatz) dahingehend signifikant voneinander unterscheiden, dass maximale Homogenität innerhalb der Gruppen und gleichzeitig maximale Heterogenität zwischen den Gruppen erreicht wird (Magidson, 1997; Schweizer & Vach, 1999). Als Segmentierungskriterium wird hierbei das empirische Signifikanzniveau (p-Wert) eines chi²- bzw. F-Testes[13] zwischen der abhängigen Variable (Behandlungsdauer) und den prädiktiven Variablen (z.B. Gesamtscore HoNOS-D) zugrunde gelegt.

Die Verfahrensgruppe stellt vier verschiedene Auswertungsalgorithmen zur Segmentierung bereit: CHAID (chi-squared automatic interaction detector), Exhaustive-CHAID, C&RT (classification and regression trees) und QUEST (quick, unbiased, efficient statistical tree) (SPSS, 2001). Diese Algorithmen unterscheiden sich z.B. hinsichtlich der zu bildenden Verzweigungen: So bildet C&RT jeweils binäre Baumknoten, während CHAID und Exhaustive-CHAID diese Vorgaben nicht haben. Für die vorliegende Arbeit wurde der Algorithmus „Exhaustive-CHAID“ (SPSS, 2001) ausgewählt, der in der Lage ist, Strukturen in höherdimensionalen Kontigenztabellen zu identifizieren. Es handelt sich hierbei um eine verfeinerte Version des CHAID-Algorithmus (Biggs et al., 1991), welcher von Kass (1980) entwickelt wurde.

---

[13] Ein chi²-Test wird berechnet, wenn die Kriteriumsvariable nicht-metrisches Skalenniveau aufweist. Der F-Test wird dementsprechend bei kontinuierlichen Kriteriumsvariablen verwendet. Ordinale Daten werden ebenfalls wie nominalskalierte Werte behandelt.

Durch die sukzessive Selektion von Prädiktoren wird eine so genannte Baumstruktur gebildet, die die Entscheidungsprozesse eines gewählten Algorithmus graphisch aufbereitet. Dabei durchläuft der Algorithmus drei Phasen, die im Folgenden näher beschrieben werden: „Merging“, „Splitting“ und „Stopping“ (Magidson, 1997).

1. *Phase „Merging“*

Das Hauptziel der Phase „Merging“ besteht darin, die Kategorien jeder Prädiktorvariablen bei der zu untersuchenden Gruppe so zu kombinieren, dass sich die neu gebildeten Kategorien in Bezug auf die Kriteriumsvariable (Behandlungsdauer) möglichst signifikant voneinander unterscheiden. Dazu werden 2-Wege Kreuztabellen in Abhängigkeit von der Kriteriumsvariablen berechnet. Für jedes Kategorienpaar, welches zusammengefasst werden soll, werden nun chi²- oder F-Werte mit den jeweils korrespondierenden p-Werten errechnet. Danach werden jene Kategorienpaare, welche sich in Bezug auf die Kriteriumsvariable (Behandlungsdauer) nicht-signifikant oder am geringsten signifikant unterscheiden, zu einer Gruppe zusammengefasst. Daraufhin werden wieder chi²- oder F-Tests der Unterkategorien der Prädiktoren in der zusammengefassten Gruppe durchgeführt. Wenn ein signifikanter p-Wert festgestellt wird, wird diese Gruppe mit der entsprechenden Merkmalskombination von der zusammengefassten Gruppe getrennt. Die „Merging-Phase“ endet für die betreffende Prädiktorvariable, wenn keine weiteren Kategorien zusammenfassbar sind (Magidson, 1997; Musiol & Steinkamp, 1998).

Die „Merging-Phase“ ist durch verschiedene Vorteile gekennzeichnet. Zum einen wird von Musiol und Steinkamp (1998) positiv hervorgehoben, dass ohne Probleme auch fehlende Werte bei den einzelnen Prädiktoren mit berücksichtigt werden können. Oft macht es auch inhaltlich Sinn, eine Zusammenfassung benachbarter Kategorien (z.B. Alter <= 29 Jahre und 30 – 39 Jahre) zuzulassen.

2. *Phase „Splitting“*

In der „Splitting-Phase“ wird nun derjenige Prädiktor mit dem niedrigsten adjustierten p-Wert (aus der Merging-Phase) identifiziert. Wenn der p-Wert kleiner ist als die maximal zulässige Irrtumswahrscheinlichkeit ($p < .05$) dann wird die jeweilige Gruppe anhand der Kategorien dieser Prädiktorvariablen in Untergruppen getrennt. Wenn kein potentieller Prädiktor einen

signifikanten p-Wert aufweist, kann die Gruppe auch nicht getrennt werden (Magidson, 1997; Musiol & Steinkamp, 1998).

3. *Phase „Stopping"*

Das Durchlaufen der ersten beiden Phasen wird für jede Prädiktorvariable solange durchgeführt, bis vorab festgelegte Abbruchkriterien erfüllt sind (Musiol & Steinkamp, 1998) („Stopping-Phase"). Diese Abbruchkriterien beziehen sich z.B. auf die Stichprobengröße des sog. Hauptknotens bzw. der Unterknoten oder auf die Verästelungstiefe des Baumes (z.B. 5 oder 10 Verästelungsebenen, Dirschedl, 1992).

Zur Definition der Abbruchkriterien finden sich in der Literatur heterogene Angaben. Während Musiol (1999) 2% des Gesamtstichprobenumfanges für den Hauptknoten und 1% des Gesamtstichprobenumfanges für die Unterknoten empfiehlt, schlägt Dirschedl (1992) vor, den Vorgang abzubrechen, wenn der Umfang der neuen Gruppen (Unterknoten) $\sqrt{N}$ unterschreitet. Zur Begrenzung der Verästelungstiefe findet sich der Hinweis von Musiol (1999), dass mit einer Erhöhung der maximalen Tiefe auch tendenziell der Prognosefehler steigt.

In der vorliegenden Arbeit wurden die Stichprobengröße des Hauptknotens ($n = 100$) und die Stichprobengröße der Subgruppen ($n = 50$) gemäß der Empfehlung von Dirschedl (1992) sowie eine maximale Baumtiefe von 5 Verästelungsebenen gemäß der Empfehlung von Musiol (1999) festgelegt. Das Signifikanzniveau für einen Prädiktor wird auf $\alpha < .05$ festgelegt.

**Zusammenfassende Bewertung des Regression-Tree-Verfahrens**

Einige Vorteile dieses Verfahrens wurden bereits genannt (z.B. Umgang mit fehlenden Werten oder Zulassen des Zusammenfügens benachbarter Kategorien). Weiterhin sind die Klassifikationsbäume in ihren Voraussetzungen relativ robust, so werden z.B. an das Skalenniveau der unabhängigen oder abhängigen Variablen keine Mindestanforderungen gestellt. Auch durch die benutzerfreundliche grafische Ergebnisdarstellung wird die Interpretation erleichtert.

Ein relativ häufig in der Literatur diskutiertes Problem stellt allerdings die Variabilität der Baumstrukturen dar. Kennzeichen eines Regression-Tree-Verfahrens ist die Auswahl von

Prädiktoren, die auf Vergleichen zwischen chi²- oder F-Werten fußen. Damit entsteht ein gewisses Maß an Willkürlichkeit in dem Fall, bei dem zwei oder mehr Prädiktoren mit annähernd gleichen chi² - oder F-Werten zur Auswahl stehen. Es wird dann der Prädiktor bevorzugt, der einen geringeren Größenunterschied aufweist. Die Wahl dieses Prädiktors führt dann – durch das rekursive Vorgehen – zur Beeinflussung der weiteren Baumstruktur. In der Arbeit von Schweizer und Vach (1999) wird deswegen zur Untersuchung der Variabilität der Baumstruktur die Nutzung des Bootstraps (Ziehen mit Zurücklegen) vorgeschlagen. Bei diesem Vorgehen werden verschiedene Bootstrap-Stichproben gezogen (ein Individuum kann in mehreren Stichproben vorkommen), für die Klassifikationsbäume bestimmt werden. Am Ende liegt eine große Zahl von Klassifikationsbäumen vor, die auch hinsichtlich der Bedeutsamkeit der einzelnen Prädiktoren untersucht werden können. Allerdings existiert bislang noch kein einheitliches Kriterium, die Vielzahl der extrahierten Baumstrukturen zu bewerten. Die Bewertung kann bislang nur auf der explorativen und deskriptiven Ebene vorgenommen werden.

Da dieses Verfahren insbesondere dazu geeignet ist, Strukturen im Datensatz zu entdecken, wird von Shepard (1995) empfohlen, die Ergebnisse in andere multivariaten Verfahren (wie z.B. der Linearen Regressionsanalyse) zu integrieren und die zusätzlich gewonnenen Informationen zu Interaktionen zwischen Prädiktoren in Bezug auf eine Kriteriumsvariable zu untersuchen. Dabei unterscheidet sich das Regression-Tree-Verfahren dahingehend von einer Linearen Regressionsanalyse, dass die Interkorrelationen zwischen den Prädiktoren, die in ein Entscheidungsbaumverfahren eingespeist werden, nicht eigens berücksichtigt werden.

#### 9.6.2.2 Weitere Auswertungsstrategien

Wenn sich die Hypothese 2a bestätigen lässt, sollen zur Überprüfung der *Hypothesen 2b und 2c* neben Regression-Tree-Verfahren auch **Multiple Regressionsanalysen** (Bortz, 1995; Barth, 1997) zur Anwendung kommen. Dabei ist der Umgang mit fehlenden Werten in der Kriteriums- und Prädiktorvariablen bei der Multiplen Regressionsanalyse von zentraler Bedeutung.

*Umgang mit fehlenden Werten in der Kriteriums- und den Prädiktorvariablen bei der Multiplen Regressionsanalyse*

Nach Barth (1997) sollte zuerst eine Kontrolle der Rohdaten erfolgen und ein Umgang mit fehlenden Werten festgelegt werden, da wechselnde Stichprobenumfänge zu einer deutlichen

Überschätzung der multiplen Korrelation R führen können. Da hier neben der Interpretation der standardisierten Regressionskoeffizienten auch die multiple Korrelation R (bzw. $R^2_{ajd}$) im Vergleich zur Varianzaufklärung des Regression-Tree-Verfahrens von Interesse ist, sollten wechselnde Stichprobenumfänge möglichst vermieden werden. Barth (1997) empfiehlt bei intervallskalierten Variablen, bei denen der Anteil fehlender Werte weniger als 10% beträgt, diese durch den Mittelwert zu ersetzen. Beträgt der Anteil allerdings über 10% sollte eine dummycodierte Variable, deren Merkmalsträger „fehlender Wert“ (1 = fehlender Wert, 0 = gültiger Wert) ist, mit in die Analyse eingespeist werden. Bei fehlenden Werten von kategorialen Daten empfehlen Little und Rubin (2002), einen EM-Algorithmus[14] einzusetzen.

Hinsichtlich der Kriteriumsvariablen wurde bereits dargestellt (vgl. Abschnitt 9.5.3), dass n = 10 Patienten, die keine gültigen Werte in der Behandlungsdauer aufwiesen, von den Analysen ausgeschlossen wurden, so dass die Variable keine fehlenden Werte mehr enthält. Eine Prüfung der Variable auf Normalverteilung ist bereits im Abschnitt 9.5.3 erfolgt.

Die Auswahl der Prädiktoren ist, wie bereits erwähnt, durch die Extraktion der Prädiktoren im Regression-Tree-Verfahren festgelegt. Die Darstellung der Verteilungseigenschaften der Merkmale findet sich in Abschnitt 9.5.1 und 9.5.2 der vorliegenden Arbeit. Eine weitere Voraussetzung der Multiplen Regressionsanalyse besteht darin, dass die Residuen, also die nicht durch die Regression erklärten Anteile des Kriteriums, unkorreliert sind. Der Durbin-Watson-Test stellt dabei die geeignete Prüfgröße dar: Ist die Testgröße hierbei $d \approx 2$, dann sind die Residuen unkorreliert. Im vorliegenden Fall nahm die Testgröße d einen Wert von 1.66 ein, was bedeutet, dass die Residuen unkorreliert und sind diese Voraussetzung somit erfüllt ist (Barth, 1997).

[14] Der EM-Algorithmus folgt dem Prinzip, dass die im Datensatz vorhandene Informationsstruktur, die aus Mittelwerten, Standardabweichungen und Korrelationen besteht, gültig ist. Auf der Basis dieser Informationen werden die fehlenden Werte zunächst geschätzt und ersetzt. Anschließend werden dann wiederum die Informationsträger (Mittelwert, SD und Korrelationen) erneut berechnet. Liegen dann Abweichungen vom Originaldatensatz vor, so kann geschlussfolgert werden, dass die ursprüngliche Informationsstruktur nicht gültig sein konnte und die neu berechnete Informationsstruktur wird beibehalten. Auf deren Basis werden dann die Kennwerte erneut geschätzt und ersetzt. Das wird so lange durchgeführt, bis eine Struktur gewonnen wurde, die sich nach der Ersetzung nicht mehr verändert (Wirtz, 2004).

In Tabelle 22 sind die Prädiktoren, die im Regression-Tree-Verfahren als signifikante Merkmale zur Vorhersage der Behandlungsdauer extrahiert wurden mit ihren Häufigkeiten gültiger und fehlender Werte aufgeführt.

Wie aus der Tabelle 22 ersichtlich, befindet sich der Prozentanteil fehlender Werte für die intervallskalierten Prädiktoren unterhalb des genannten 10%-Kriteriums. Es wurde entsprechend den Empfehlungen von Barth eine Mittelwertsersetzung vorgenommen (Barth, 1997). Bei den kategorialen Merkmalen befindet sich der Prozentanteil fehlender Werte für die Prädiktoren „Partnersituation“, „ambulante Psychotherapie“ und „Komorbidität“ oberhalb des definierten 10%-Kriteriums.

Im nächsten Schritt wurde deswegen empirisch mittels t-Tests überprüft, ob es einen systematischen Zusammenhang zwischen dem Auftreten der fehlenden Werte in diesen Prädiktoren und der Kriteriumsvariable „Behandlungsdauer“ gibt (vgl. Tabelle 23).

Tabelle 22: Häufigkeiten fehlender Werte der Prädiktoren des Fallgruppenbaumes (Variablen auf Intervallskalen- und Nominalskalenniveau)

| | **N** | **Fehlend** | |
|---|---|---|---|
| | | **N** | **%** |
| **Intervallskalenniveau** | | | |
| Schweregradzusatz der Erstdiagnose | 1816 | 91 | 4.8 |
| Alter | 1835 | 72 | 3.8 |
| SF-8-S (psych. S.) | 1878 | 29 | 1.5 |
| IIP-25 (kalt/nicht unterst.) | 1814 | 93 | 4.9 |
| SF-8-F (psych. S.) | 1893 | 14 | 0.7 |
| HoNOS-D (12) | 1897 | 10 | 0.5 |
| BMI | 1790 | 117 | 6.1 |
| Chronifizierung | 1860 | 47 | 2.5 |
| **Nominalskalenniveau** | | | |
| **Partnersituation** | **1634** | **273** | **14.3** |
| Erstdiagnose | 1897 | 10 | 0 |
| **Ambulante Psychotherapie** | **1649** | **258** | **13.5** |
| **Komorbidität** | **1375** | **532** | **27.9** |

**Anmerkungen:** SF-8-S = Short-Form-8-Selbsteinschätzung; IIP-25 = Kurzversion des Inventars Interpersoneller Probleme; SF-8-F = Short-Form-8-Fremdeinschätzung; HoNOS-D (12) = Gesamtscore der HoNOS-D über alle 12 Items; BMI = Body Mass Index

Es zeigen sich keine signifikanten Zusammenhänge zwischen dem Nicht-Beantworten der Variable „Partnersituation“ und der Behandlungsdauer. Demgegenüber unterscheiden sich die Patienten mit fehlenden Werten in den Merkmalen „Komorbidität“ und „ambulante Psychotherapie“ hinsichtlich der Behandlungsdauer signifikant mit annähernd kleiner Effektstärke von den Patienten mit gültigen Werten (vgl. Tabelle 23).

Da in dieser Analyse systematische Unterschiede zwischen dem Nicht-Beantworten und dem Beantworten der Merkmale bezüglich der Behandlungsdauer festgestellt wurden, empfiehlt sich nach Barth (1997) die fehlenden Werte als eine separate dummycodierte Variable zu modellieren und in die Lineare Regressionsanalyse mit einzuspeisen. Zusätzlich wurden die fehlenden Werte der einzelnen Merkmale „Partnersituation“, „Komorbidität“ und „ambulante Psychotherapie“ mittels des empfohlenen EM-Algorithmus (Little & Rubin, 2002) unter Berücksichtigung der Werte der anderen Prädiktoren und der Behandlungsdauer ersetzt. Dieses Vorgehen wurde gewählt, um den Einfluss systematischer Variationen auf die Behandlungsdauer zu minimieren.

Tabelle 23: Vergleich der Behandlungsdauer zwischen „Partnersituation“, „Komorbidität“ und „ambulante Psychotherapie“ gültige vs. fehlende Werte

| | Behandlungsdauer | | | |
|---|---|---|---|---|
| | n | M | SD | t-Wert (df), p-Wert, d |
| **Partnersituation** | | | | |
| Fehlender Wert | 273 | 47.19 | 16.29 | $t_{(df = 1905)}{}^{1} = -.31$, $p = .75$, $d = .02$ |
| Alle anderen Kategorien | 1634 | 47.55 | 17.60 | |
| **Komorbidität** | | | | |
| Fehlender Wert | 532 | 49.87 | 20.14 | $t_{(df = 808.87)}{}^{2} = 3.37$, $p \leq .001$, $d = .19$ |
| Alle anderen Kategorien | 1375 | 46.58 | 16.16 | |
| **Ambulante Psychotherapie** | | | | |
| Fehlender Wert | 258 | 49.83 | 16.66 | $t_{(df = 1905)}{}^{1} = 2.31$, $p = .02$, $d = .16$ |
| Alle anderen Kategorien | 1649 | 47.14 | 17.51 | |

**Anmerkungen:** Die Effektstärke d wurde anhand der gepoolten Standardabweichung berechnet (Cohen, 1988)[15], Verwendet wurde jeweils der Betrag von d; [1] = t-Test bei homogenen Varianzen (Levene-Test); [2] = t-Test bei inhomogenen Varianzen (Levene-Test)

Nach der Berechnung der Multiplen Regressionsanalysen sollen die Prädiktoren „Gesamtscore der HoNOS-D“ und „Psychische Summenskala der SF-8-F“ entsprechend der Hypothese 2b und 2c bewertet werden. Dazu ist eine Auswertungsstrategie erforderlich, die nachfolgend erläutert werden soll.

*Auswertungsstrategie zur Beurteilung der Bedeutsamkeit der Prädiktoren bei der Multiplen Regressionsanalyse*

Um zu einer vergleichbaren Aussage über die Bedeutsamkeit der Prädiktoren „Gesamtscore der HoNOS-D“ und „Psychische Summenskala der SF-8-F“ zu gelangen, sollen ihre aus der Multiplen Regressionsanalyse gewonnenen Regressionskoeffizienten gegeneinander getestet

---

[15] Folgende Formel wurde zur Berechnung der gepoolten Standardabweichung zugrunde gelegt:

$$SD_{gepoolt} = \sqrt{\frac{(n_1 - 1) * s_1{}^2 + (n_2 - 1) * s_2{}^2}{n_1 + n_2 - 2}}$$ (Cohen, 1988)

werden. Um dabei eine direkte Vergleichbarkeit der Regressionskoeffizienten zu gewährleisten, werden die Skalenwerte *z-standardisiert* in die Multiple Regressionsanalyse eingespeist.

Unter Annahme der Student-t-Verteilung werden dann die unstandardisierten Regressionskoeffizienten der beiden Prädiktoren anhand ihrer Standardmessfehler gegeneinander getestet (Bortz & Döring, 1995). Die Standardmessfehler werden im Rahmen der Multiplen Regressionsanalyse als Kennwert der Koeffizienten ausgegeben. Als statistische Prüfgrößen werden die empirische und die kritische Differenz der beiden Regressionskoeffizienten (z.B. $B_{SF\text{-}8\text{-}F} - B_{HoNOS\text{-}D}$) berechnet (Amelang & Zielinski, 1997). Getestet werden soll auf einem α-Signifikanzniveau von .05.

Folgende Formel wird zur Berechnung der kritischen Differenz zugrunde gelegt:

$Diff_{krit} = z_{\alpha/2} * s_{me\text{-}Diff}$ (Amelang & Zielinski, 1997) (Formel 2.32, S. 54)

**$Diff_{krit}$** = kritische Differenz

**$z_{\alpha/2}$** = z-Wert bei entsprechendem α -Niveau (für α = .05 ergibt sich ein z-Wert von 1.96, Bortz & Döring, 1995)

**$s_{me\text{-}Diff}$** = Messfehler der Differenz.

Zur Bestimmung der kritischen Differenz benötigt man den Messfehler der Differenz (**$s_{me\text{-}Diff}$**), der sich anhand folgender Formel aus den Standardmessfehlern der einzelnen Koeffizienten bestimmen lässt.

$s_{me-Diff} = \sqrt{s_{me1}^2 + s_{me2}^2}$ (Amelang & Zielinski, 1997, Formel 2.30, S. 54)

$s_{me1}$, $s_{me2}$ = Standardmessfehler der Regressionskoeffizienten.

# 10 Ergebnisse

Zunächst sollen die Ergebnisse zur generellen Eignung der HoNOS-D (Hypothese 1) dargestellt werden. Danach folgen die Resultate zur speziellen Eignung der HoNOS-D (Hypothese 2) im Rahmen der Fallgruppenbildung.

## 10.1 Die generelle Eignung der HoNOS-D (Fragestellung 1)

### *10.1.1 Analysen zur Praktikabilität (Hypothese 1a)*

Eine deskriptive Analyse zur Anzahl fehlender Werte ergab für 11 der 12 Items zufrieden stellende Ergebnisse mit einer durchschnittlichen Rate von 3.3%[16]. Die Spannweite reichte für 11 der 12 Items von 1.3% bis 4.5%. So wurde z.B. das Item 7 „Gedrückte Stimmung“ nur in 1.3% der Fälle nicht angegeben. Auch für den „Gesamtscore der HoNOS-D“ zeigt sich mit 3.4% ein befriedigender prozentualer Anteil fehlender Werte (siehe Tabelle 24). Für die Berechnung des Gesamtscores sollten mindestens 10 der 12 Items beantwortet sein.

Erwartungsgemäß wies Item 8 „Andere Symptome“ mit 8.6% die höchste Prozentzahl fehlender Werte auf (siehe Tabelle 24). Überraschenderweise zeigte sich für die Items 11 und 12 ein prozentualer Anteil fehlender Werte, der deutlich unter 10% liegt (4.5% und 3.8%) (siehe Tabelle 24).

Im Rahmen einer Schulung von Therapeuten verschiedener Berufsgruppen (Ärzte, Psychologen) (n = 7), die die HoNOS-D zur Beurteilung suchtkranker Patienten in stationärer Rehabilitationsbehandlung einsetzten, wurden Fragen zur Praktikabilität der HoNOS-D gestellt (siehe Fragebogen im Anhang 16.4.1). Angesichts des vergleichsweise geringen Stichprobenum-

---

[16] Aufgrund fehlender Werte umfasst die Stichprobengröße in den nachfolgenden Analysen weniger als N = 1.918 Patienten.

fangs können die Ergebnisse lediglich als Ergänzung zu den oben aufgeführten Analysen bewertet werden.

Tabelle 24: Verteilungskennwerte der 12 HoNOS-D Items

| Item<br>Nummer/Bezeichnung | Missings (%) | M | SD | Median | Schiefe | Kurtosis | α |
|---|---|---|---|---|---|---|---|
| 1. Aggression | 3.2 | 0.74 | 1.13 | 0 | 1.27 | 0.37 | |
| 2. Selbstverletzung | 3.4 | 0.21 | 0.72 | 0 | 3.66 | 12.80 | |
| 3. Substanzmissbrauch | 3.9 | 0.26 | 0.76 | 0 | 3.01 | 8.34 | |
| 4. Kognitive Probleme | 2.0 | 1.22 | 1.29 | 1 | 0.54 | -1.01 | |
| 5. Probleme durch körperliche Erkrankungen | 2.3 | 1.21 | 1.34 | 1 | 0.59 | -1.04 | |
| 6.Halluzinationen und Wahnvorstellungen | 2.9 | 0.11 | 0.52 | 0 | 5.26 | 29.01 | |
| 7. Depression | 1.3 | 2.80 | 0.85 | 3 | -0.78 | 1.15 | |
| 8. Andere Symptome | 8.6 | 2.97 | 0.97 | 3 | -1.40 | 2.36 | |
| 9. Probleme mit Beziehungen | 2.2 | 2.56 | 1.11 | 3 | -0.82 | 0.16 | |
| 10. Probleme durch alltägliche Aktivitäten | 2.2 | 1.84 | 1.16 | 2 | -0.24 | -0.89 | |
| 11. Probleme durch die Wohnbedingungen | 4.5 | 0.70 | 1.10 | 0 | 1.38 | 0.66 | |
| 12. Probleme im Beruf und in der Freizeit | 3.8 | 1.94 | 1.36 | 2 | -0.21 | -1.23 | |
| **Gesamtscore der HoNOS-D** | **3.4** | **1.37** | **0.48** | | **0.32** | **0.05** | **.64** |

**Anmerkungen:** M = Mittelwert; SD = Standardabweichung; α = Cronbach´s Alpha.

Eine deskriptive Analyse der Daten aus der Praktikabilitätsbefragung ergab, dass die im HoNOS-D bereits erfahrenen Beurteiler im Mittel 5 Minuten zum Ausfüllen des Instrumentes benötigten. Dabei nannten die Therapeuten am häufigsten Probleme mit dem Item 12 (5 Nennungen), gefolgt vom Item 8 (3 Nennungen) und den Items 10 und 11 (jeweils 2 Nennungen). Die Verständlichkeit des Glossars wurde auf einer Skala von 0 = sehr schlecht bis 5 = ausgezeichnet im Durchschnitt mit einem befriedigenden Wert von 3.1 eingeschätzt. Des Weiteren führte die Mehrzahl der Beurteiler an, alle nötigen Informationen für die Einschätzung im HoNOS-D aus den routinemäßig durchgeführten Gesprächen zu erhalten. Eine detaillierte Auflistung der Ergebnisse dieser eher explorativen Analyse findet sich im Anhang 16.4.2.

### *10.1.2 Analysen zur Reliabilität der HoNOS-D*

#### 10.1.2.1 Verteilungscharakteristika der HoNOS-D Items (Hypothese 1b)

Die Verteilungseigenschaften der 12 HoNOS-D Items sind in Abbildung 42, die dazugehörigen Kennwerte in Tabelle 24 dargestellt. In der zugrunde liegenden Stichprobe zeigte sich erwartungsgemäß, dass Item 3 „Substanzmissbrauch" und Item 6 „Halluzinationen und Wahnvorstellungen" extrem rechtsschiefe Verteilungsformen aufweisen, wobei der Wert 0 (= kein Problem) in 84% bzw. 91% der Fälle vergeben wurde. Entgegen der Hypothese 1b zeigte sich weiterhin, dass auch Item 2 „Selbstverletzung" deutlich rechtsschief verteilt war. Des Weiteren ergab sich für die Items 1, 5 und 11 ein limitierter Streubereich: Am häufigsten wurde der Wert 0 vergeben und 95% der Fälle liegen in einem Bereich zwischen 0 und 2 (Item 1 „Aggression", Item 5 „Probleme mit körperlichen Erkrankungen") bzw. 0 bis 1 (Item 11 „Probleme durch die Wohnbedingungen"). Ausreichende Variation hingegen weisen die anderen 6 der 12 Items auf: Item 4 „Kognitive Probleme", Item 7 „Gedrückte Stimmung", Item 8 „Andere Symptome", Item 9 „Probleme mit Beziehungen", Item 10 „Probleme mit alltäglichen Aktivitäten" und Item 12 „Probleme im Beruf und in der Freizeit" (siehe Abbildung 42). Ihre Schiefeindizes umfassen eine Spanne zwischen – 1.4 bis + 1.4 (siehe Tabelle 24).

Für den „Gesamtscore der HoNOS-D" finden sich mit einer Schiefe von 0.32 und einer Kurtosis von 0.05 zufrieden stellende Verteilungseigenschaften (siehe Abbildung 43 und Tabelle 24). Dagegen befindet sich Cronbach´s alpha mit .64 in einem eher moderaten Bereich.

Die höchsten mittleren Ausprägungen im Schweregrad ($M \geq 2.50$) sind in den Items 7 „Depressive Stimmung", Item 8 „Andere Symptome" und im Item 9 „Probleme mit Beziehungen" zu finden (siehe Tabelle 24).

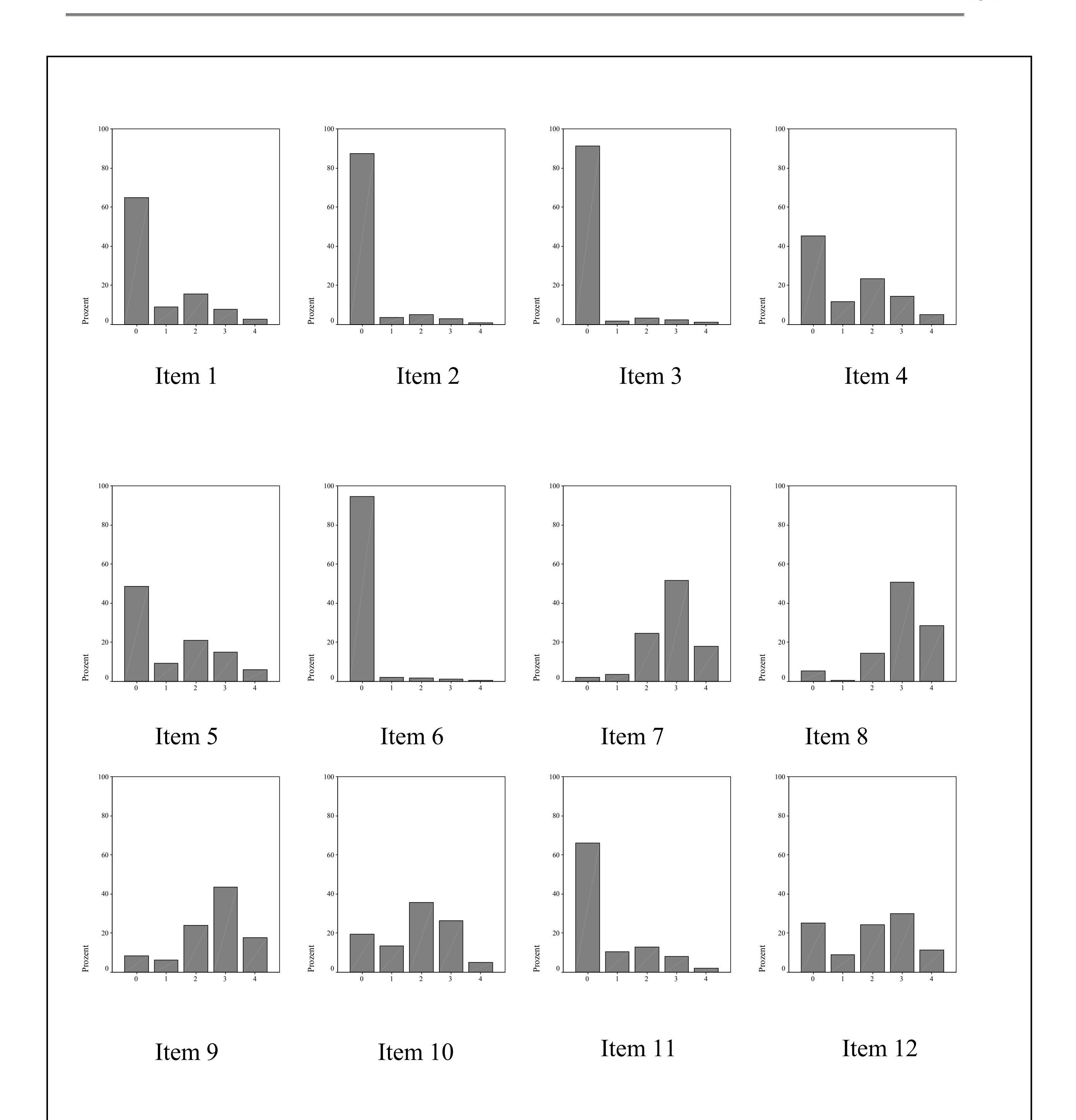



Abbildung 42: Itemverteilungen der HoNOS-D in Prozent (Kategorien 0 bis 4)

**Anmerkungen:** Item 1 = Aggression; Item 2 = Selbstverletzung; Item 3 = Substanzmissbrauch; Item 4 = Kognitive Probleme; Item 5 = Probleme mit körperlichen Erkrankungen; Item 6 = Halluzinationen und Wahnvorstellungen; Item 7 = Gedrückte Stimmung; Item 8 = Andere Symptome; Item 9 = Probleme mit Beziehungen; Item 10 = Probleme mit alltäglichen Aktivitäten; Item 11 = Probleme durch die Wohnbedingungen; Item 12 = Probleme im Beruf und in der Freizeit

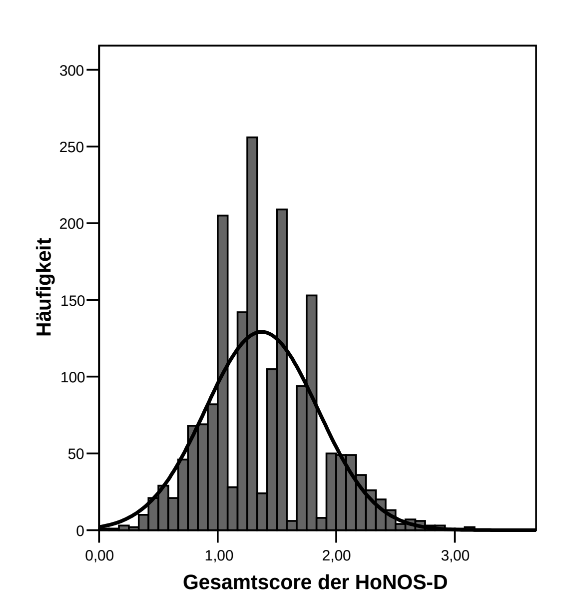



Abbildung 43: Verteilungseigenschaften des „Gesamtscores der HoNOS-D“ (Angaben in Häufigkeiten)

#### 10.1.2.2 Item-Interkorrelationen (Hypothese 1c)

Die Testautoren intendierten bei der Skalenkonstruktion, dass jedes Item für sich genommen eine ein distinktes Merkmal sein sollte und dementsprechend die Korrelationen zwischen den Items gering sein sollten (Wing et al., 1996). Die Analyse der Iteminterkorrelationsmatrix zeigt weitestgehend zufrieden stellende Ergebnisse. Erwartungskonform wies nur ein Itempaar eine signifikante Korrelation mit mittlerer Effektstärke über $r > .30$ auf: Item 10 „Probleme mit alltäglichen Aktivitäten“ hing mit dem Item 7 „Gedrückte Stimmung“ zusammen. Keine andere Korrelation erreichte die Höhe einer mittleren Effektstärke oder eines Zusammenhanges von über .40 (siehe Tabelle 25). Allerdings zeigten sich signifikante Zusammenhänge mindestens kleiner Effektstärke ($r > .27$) zwischen Item 4 „Kognitive Probleme“ und Item 10 „Probleme mit alltäglichen Aktivitäten“, zwischen Item 9 „Probleme mit Beziehungen“ und Item 10 „Probleme mit alltäglichen Aktivitäten“, zwischen Item 9 „Probleme mit Beziehungen“ und Item 11 „Probleme durch die Wohnbedingungen“ und zwischen Item 10 „Probleme mit alltäglichen Aktivitäten“ und Item 12 „Probleme im Beruf und in der Freizeit“ (siehe Tabelle 25).

Tabelle 25: Interkorrelationsmatrix der HoNOS-D Items zu Beginn der Behandlung (Spearmann-Rho Korrelationen) (N = 1918)

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 | Item 8 | Item 9 | Item 10 | Item 11 | Item 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item 1 | 1.00 | | | | | | | | | | | |
| Item 2 | .235** | 1.00 | | | | | | | | | | |
| Item 3 | .151** | .155** | 1.00 | | | | | | | | | |
| Item 4 | .197** | .101** | .100** | 1.00 | | | | | | | | |
| Item 5 | .076** | -.038 | -.054* | .158** | 1.00 | | | | | | | |
| Item 6 | .057** | -.023 | .107** | .125** | -.046* | 1.00 | | | | | | |
| Item 7 | -.010 | .048* | .021 | .141** | .060* | -.039 | 1.00 | | | | | |
| Item 8 | .141** | .160** | .023 | .149** | .066** | -.027 | .263** | 1.00 | | | | |
| Item 9 | .190** | .173** | .074** | .132** | -.048* | .064** | .270** | .240** | 1.00 | | | |
| Item 10 | .095** | .112** | .058* | .292** | .195** | .076** | **.358**** | .265** | .276** | 1.00 | | |
| Item 11 | .160** | .161** | .069** | .182** | .027 | .067** | .124** | .146** | .297** | .252** | 1.00 | |
| Item 12 | .090** | .027 | .047* | .155** | .091** | .018 | .209** | .116** | .112** | .297** | .215** | 1.00 |

**Anmerkungen:** Item 1 = Aggression; Item 2 = Selbstverletzung; Item 3 = Substanzmissbrauch; Item 4 = Kognitive Probleme; Item 5 = Probleme mit körperlichen Erkrankungen; Item 6 = Halluzinationen und Wahnvorstellungen; Item 7 = Gedrückte Stimmung; Item 8 = Andere Symptome; Item 9 = Probleme mit Beziehungen; Item 10 = Probleme mit alltäglichen Aktivitäten; Item 11 = Probleme durch die Wohnbedingungen; Item 12 = Probleme im Beruf und in der Freizeit

#### 10.1.2.3 Interrater-Reliabilitäten (Hypothese 1d)

In zwei ergänzend erhobenen Stichproben schätzten jeweils zwei Beurteiler die Patienten in den Items der HoNOS-D ein. Die Ergebnisse der Analysen zur Interrater-Reliabilitäten sind in den beiden nachfolgenden Tabellen zusammengetragen (siehe Tabelle 26 und Tabelle 27).

Für die beurteilten Patienten aus stationärer Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie (N = 37[17]) befinden sich die Gesamtwerte der Intraklassenkorrelationen (ICC) zwischen .00 und .78, für die Patienten der stationären Rehabilitationseinrichtung Suchtkranker (N = 34) liegen die Werte in einem Bereich von .00 bis .55 (vgl. Tabelle 26 und Tabelle 27). Die Finn-Koeffizienten nahmen für die stationäre Krankenhausbehandlung

[17] Eine Kalkulation der Power mit dem SPSS Programm Sample Power 2.0 ergab, dass bei gegebenen N = 37 Patienten und einem befriedigendem Schwellenwert des Intraklassenkorrelationskoeffizienten von .70, die erzielten Werte der Interraterreliabilitäten zwischen .37 und .70 einer hohen statistischen Unsicherheit unterliegen. Nur für die Werte unterhalb von .37 (ICC) kann davon ausgegangen werden, dass sie sich in anderen Studien mit 80%-iger Wahrscheinlichkeit replizieren lassen. Zur Absicherung des Unsicherheitsbereiches zwischen .37 und .70 bedarf es also größerer Stichprobenumfänge.

im Bereich Psychosomatik und Psychotherapie Werte zwischen .28 und .90 an, für die Patienten der stationären Rehabilitationseinrichtung Suchtkranker liegen die Koeffizienten in einem Bereich von .83 bis .96.

Die Beurteilerübereinstimmung[18] aus der Teilstichprobe der stationären Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie befinden sich für die Items 1 „Aggression“, 3 „Substanzmissbrauch“, 4 „ Kognitive Probleme“, 6 „Halluzinationen und Wahnvorstellungen“, 7 „Gedrückte Stimmung“, 9 „Probleme mit Beziehungen“ und 10 „Probleme mit alltäglichen Aktivitäten“ in einem akzeptablen Bereich (siehe Tabelle 26). Der Finn-Koeffizient für den Gesamtscore der HoNOS-D weist für die Teilstichprobe 1 insgesamt eine befriedigende Höhe auf (Finn > .70) (siehe Tabelle 26).

[18] Es wurden in der vorliegenden Arbeit sowohl die ICC als auch die prozentuale Übereinstimmung und die Finn-Koeffizienten berechnet werden. Die Beurteilung der Güte der Interrater-Reliabilitäten wurde wie folgt festgelegt: Sobald ein Wert der drei berichteten Koeffizienten innerhalb des akzeptablen Bereiches liegt, wird eine ausreichende Interrater-Reliabilität für das Item der HoNOS-D angenommen (vgl. hierzu auch Abschnitt 9.6.1.1).

Tabelle 26: Interkorrelationskoeffizienten zur Bestimmung der Interrater-Reliabilität von N = 37 Patienten aus stationärer psychosomatischer/psychotherapeutischer Krankenhausbehandlung

| Items der HoNOS-D | **Raterpaar 1** ICC (n = 4) (%-tuale Überein-stimmung) | **Raterpaar 2** ICC (n = 15) (%-tuale Überein-stimmung) | **Raterpaar 3** ICC (n = 11) (%-tuale Überein-stimmung) | **Raterpaar 4** ICC (n = 7) (%-tuale Über-einstimmung) | ICC Gesamt (Fisher z-transformiert) n = 37 | Finn-Koeffizient (Gesamt n = 37) |
|---|---|---|---|---|---|---|
| 1. Aggression | 1.00 (100%) | 0.455 (60%) | 0.762 (90%) | 0.341 (71%) | .69 | .89 |
| **2. Selbstverletzung** | **n.b. (100%)** | **n.b. (66%)** | **0.036 (36%)** | **0.015 (14%)** | **.02** | **.52** |
| 3. Substanzmissbrauch | n.b. (75%) | n.b. (50%) | .826 (54%) | .692 (42%) | .78 | .91 |
| 4. Kognitive Probleme | n.b. (75%) | n.b. (80%) | 0.00 (63%) | n.b. (85%) | .00 | .90 |
| **5. Probleme mit körperlichen Erkrankungen** | **n.b. (50%)** | **n.b. (73%)** | **.350 (54%)** | **0.00 (28%)** | **.22** | **.28** |
| 6. Halluzinationen und Wahnvorstellungen | n.b. (100%) | n.b. (93%) | .643 (91%) | .143 (57%) | .48 | .32 |
| 7. Gedrückte Stimmung | n.b. (50%) | n.b. (66%) | .130 (63%) | .486 (57%) | .27 | .87 |
| **8. Andere Symptome** | **n.b. (25%)** | **n.b. (60%)** | **.479 (54%)** | **0.00 (42%)** | **.30** | **.62** |
| 9. Probleme mit Beziehungen | n.b. (0%) | n.b. (40%) | .699 (54%) | .368 (42%) | .58 | .85 |
| 10. Probleme mit alltäglichen Aktivitäten | n.b. (75%) | n.b. (46%) | .496 (36%) | .357 (28%) | .44 | .85 |
| **11. Probleme durch die Wohnbedingungen** | **n.b. (50%)** | **n.b. (47%)** | **.130 (64%)** | **n.b. (43%)** | **.13** | **.67** |
| **12. Probleme im Beruf und in der Freizeit** | **n.b. (50%)** | **n.b. (27%)** | **.107 (54%)** | **n.b. (29%)** | **.10** | **.66** |
| **Gesamtscore** | | | | | .33 | **.70** |

**Anmerkung:** ICC = Intraklassenkoeffizient;%-tuale Übereinstimmung = Prozentuale Übereinstimmung; n.b. = aufgrund fehlender Variation wurde der ICC *n*icht *b*erechnet. Die Varianzen innerhalb der Gruppen, die der Berechnung des Finn-Koeffizienten zugrunde liegen befinden sich im Anhang 16.5.1.; fett markiert = moderate Werte in der Beurteilerübereinstimmung

Die Beurteilungen der Patienten aus der stationären Rehabilitationseinrichtung Suchtkranker dagegen weisen für alle 12 Items der HoNOS-D akzeptable Finn-Koeffizienten auf (siehe Tabelle 27). Auch für die Teilstichprobe 2 zeigt sich mit .83 für den „HoNOS-D Gesamtscore“ ein zufrieden stellender Betrag des Finn-Koeffizienten (siehe Tabelle 27).

Tabelle 27: Interkorrelationskoeffizienten zur Bestimmung der Interrater-Reliabilität von N = 34 Patienten einer stationären Rehabilitationsbehandlung Suchtkranker

| Items der HoNOS-D | ICC Gesamt | Prozentuale Übereinstimmung | Finn-Koeffizient (Gesamt n = 34) |
|---|---|---|---|
| 1. Aggression | .20 | 56% | .91 |
| 2. Selbstverletzung | .55 | 94% | .85 |
| 3. Substanzmissbrauch | n.b. | 82% | .96 |
| 4. Kognitive Probleme | .00 | 47% | .89 |
| 5. Probleme mit körperlichen Erkrankungen | .33 | 30% | .85 |
| 6. Halluzinationen und Wahnvorstellungen | n.b. | 88% | .94 |
| 7. Gedrückte Stimmung | .45 | 59% | .92 |
| 8. Andere Symptome | .35 | 59% | .90 |
| 9. Probleme mit Beziehungen | .30 | 68% | .92 |
| 10. Probleme mit alltäglichen Aktivitäten | .38 | 50% | .92 |
| 11. Probleme durch die Wohnbedingungen | .42 | 53% | .85 |
| 12. Probleme im Beruf und in der Freizeit | .00 | 45% | .85 |
| **Gesamtscore** | | | **.83** |

**Anmerkungen:** ICC = Intraklassenkorrelationen; n.b. = aufgrund fehlender Variation wurde der ICC nicht berechnet. Die Varianzen innerhalb der Gruppen, die der Berechnung des Finn-Koeffizienten zugrunde liegen befinden sich im Anhang 16.5.1.

#### 10.1.2.4 Test- Retest Reliabilität (Hypothese 1e)

Die Ergebnisse zur Test-Retest Reliabilität für die Subgruppe derjenigen Patienten, die sich während der stationären Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen in zentralen Outcomeparametern nicht veränderten, sind in Tabelle 28 dargestellt. Es zeigten sich für 9 der 12 Items und den Gesamtscore der HoNOS-D Intraklassenkoeffizienten in einem zufrieden stellenden Bereich von .80 bis .92. Für drei der 12 Items hingegen (Item 3 „Substanzmissbrauch“, Item 7 „Gedrückte Stimmung“ und Item 8 „Andere Symptome“) zeigten sich zwar noch befriedigende Werte zwischen .72 und .75, die aber etwas unterhalb der Höhe der Koeffizienten für die anderen Items lagen (vergleiche Tabelle 28).

Tabelle 28: Intraklassenkoeffizienten zur Bestimmung der Test-Retest-Reliabilität der HoNOS-D (N = 540)

| Items der HoNOS-D | ICC (prä-post) |
|---|---|
| 1. Aggression | .80 |
| 2. Selbstverletzung | .87 |
| 3. Substanzmissbrauch | .73 |
| 4. Kognitive Probleme | .89 |
| 5. Probleme mit körperlichen Erkrankungen | .92 |
| 6.Halluzinationen und Wahnvorstellungen | .91 |
| 7. Gedrückte Stimmung | .75 |
| 8. Andere Symptome | .72 |
| 9. Probleme mit Beziehungen | .84 |
| 10. Probleme mit alltäglichen Aktivitäten | .83 |
| 11. Probleme durch die Wohnbedingungen | .90 |
| 12. Probleme im Beruf und in der Freizeit | .89 |
| **Gesamtscore** | **.91** |

**Anmerkung:** ICC = Intraklassenkorrelation

### 10.2 Zusammenfassung der Ergebnisse zur generellen Eignung der HoNOS-D (Hypothese 1a bis 1e)

Im Folgenden sollen die Hypothesen noch einmal in zusammengefasster Form aufgeführt und hinsichtlich ihrer Bestätigung bzw. Verwerfung bewertet werden. Einen Gesamtüberblick über die Ergebnisse für die Prüfung der Hypothesen 1a bis 1e vermittelt die Tabelle 29.

**Hypothese 1a „Praktikabilität“:** Es wird erwartet, dass 9 der 12 Items und der Gesamtscore der HoNOS-D fehlende Werte unter 10% aufweisen. Für die Items 8, 11 und 12 werden fehlende Werte über 10% erwartet.

Für die Items 1 bis 7, 9 und 10 der HoNOS-D und den „Gesamtscore“ kann die Hypothese 1a aufgrund eines prozentualen Anteils fehlender Werte unter 10% bestätigt werden. Für Item 8, 11 und 12 der HoNOS-D wird die Hypothese 1a aufgrund des prozentualen Anteils fehlender Werte unter 10% verworfen.

**Hypothese 1b „Verteilungseigenschaften“:** Für die Items 3 und 6 der HoNOS-D werden rechts schiefe Verteilungseigenschaften erwartet. Alle anderen Items und der „Gesamtscore der HoNOS-D“ sollten moderate bis zufrieden stellende Verteilungseigenschaften aufweisen.

Die Hypothese 1b kann für die Items 3 und 6 aufgrund rechts schiefer Verteilungseigenschaften bestätigt werden. Die Hypothese 1b kann weiterhin für die Items 1, 4, 5, 7, 8, 9, 10, 11 und 12 der HoNOS-D sowie für den „Gesamtscore der HoNOS-D" aufgrund moderater bzw. zufrieden stellender Verteilungseigenschaften (Schiefeindizes -1.4 bis +1.4) bestätigt werden. Die Hypothese 1b wird aufgrund deutlich rechts schiefer Verteilungseigenschaften für das Item 2 der HoNOS-D verworfen.

**Hypothese 1c „Interkorrelation":** Es werden höchstens zwei signifikante Zusammenhänge mindestens mittlerer Effektstärke ($r > .40$) zwischen den Items der HoNOS-D erwartet. Es können signifikante Zusammenhänge zwischen dem Item 4 und Item 10 sowie zwischen Item 2 und Item 7 erwartet werden.

Die Hypothese 1c kann aufgrund nur einer signifikanten Korrelation mit mittlerer Effektstärke zwischen Item 10 und Item 7 ($r = .358$) bestätigt werden. Alle anderen Item-Interkorrelationen erreichen höchstens kleine Effektstärken. Die Hypothese weiterhin postulierten Zusammenhänge konnten nicht aufgezeigt werden, dagegen korrelierte mit mindestens mittlerer Effektstärke das Item 7 „Gedrückte Stimmung" mit dem Item 10 „Probleme mit alltäglichen Aktivitäten".

**Hypothese 1d „Interraterreliabilität":** Für die Items 4, 8, 9, 10, 11 und 12 der HoNOS-D werden niedrigere Interrater-Übereinstimmungen (Finn-Koeffizient oder ICC unter < .70 oder prozentuale Übereinstimmung < 50%) erwartet. Für die Items 1, 2, 3, 5, 6 und 7 sowie den „Gesamtscore der HoNOS-D" werden befriedigende Werte (Koeffizienten über > .70 bzw. über 50%) erwartet.

Diese Hypothese soll hinsichtlich der beiden unterschiedlichen untersuchten Teilstichproben differenziert betrachtet werden.

*Teilstichprobe 1 (Patienten aus stationärer psychosomatischer/psychotherapeutischer Krankenhausbehandlung)*: Die Hypothese 1d kann für die Items 1, 3, 6, 7 und für den „Gesamtscore der HoNOS-D" aufgrund befriedigender Werte in den ermittelten Koeffizienten bestätigt werden. Für die Items 8, 11 und 12 der HoNOS-D kann die Hypothese 1d aufgrund niedriger Übereinstimmungsraten ebenfalls bestätigt werden. Für die übrigen Items der HoNOS-D muss

die Hypothese 1d entweder aufgrund niedrigerer als erwarteter Koeffizienten (Item 2 und 5) bzw. höherer als erwarteter Koeffizienten (Item 4, 9 und 10) verworfen werden.

*Teilstichprobe 2 (Patienten aus stationärer Rehabilitationsbehandlung Suchtkranker)*: Die Hypothese 1d kann für die Items 1, 2, 3, 5, 6 und 7 sowie für den „Gesamtscore der HoNOS-D“ aufgrund befriedigender Werte der ermittelten Koeffizienten bestätigt werden. Für die übrigen Items (4, 8, 9, 10, 11 und 12) der HoNOS-D muss die Hypothese 1d aufgrund höherer Übereinstimmung als erwartet verworfen werden.

*Zusammenfassend* lässt sich für die beiden untersuchten Teilstichproben aussagen, dass für die Items 1, 3, 6, 7 und den „Gesamtscore der HoNOS-D“ die Koeffizienten im erwarteten Bereich lagen.

**Hypothese 1e „Test-Retest-Reliabilität“:** Für 8 der 12 Items sowie den „Gesamtscore der HoNOS-D“ werden befriedigende Werte der Interklassenkoeffizienten (über > .70) erwartet. Für Item 1, 3, 7 und 10 der HoNOS-D werden niedrigere Werte der Test-Retest-Koeffizienten (unter < .70) erwartet.

Für die Items 2, 4, 5, 6, 8, 9, 11 und 12 sowie den „Gesamtscore der HoNOS-D“ kann die Hypothese 1e aufgrund von Koeffizienten über .70 bestätigt werden. Die Hypothese 1e muss dagegen für die Items 1, 3, 7 und 10 der HoNOS-D aufgrund höherer als erwarteter Werte verworfen werden.

Auf der Basis der Analysen zur generellen Eignung der HoNOS-D kann geschlussfolgert werden, dass der **„Gesamtscore der HoNOS-D“**, welcher über alle 12 Items gebildet wird (Voraussetzung: mindestens 10 Items beantwortet), aufgrund inhaltlicher Überlegungen und aufgrund der moderaten faktoriellen Struktur der Subskalen als Maß für die Schweregradbelastung in die Fallgruppenbildung eingehen soll.

Tabelle 29: Differenzierte Auflistung der Bestätigung bzw. Verwerfung der Hypothesen 1a bis 1e pro Item der HoNOS-D

| **Hypothesen** / **Items der HoNOS-D** | **1a** | **1b** | **1c** | **1d** | | **1e** |
|---|---|---|---|---|---|---|
| | | | | Teilstichprobe 1 | Teilstichprobe 2 | |
| Item 1 „Aggression“ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ |
| Item 2 „Selbstverletzung“ | ✓ | ✗ | ✗ | ✗ | ✓ | ✓ |
| Item 3 „Substanzmissbrauch“ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ |
| Item 4 „Kognitive Probleme“ | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ |
| Item 5 „Probleme mit körperlichen Erkrankungen“ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ |
| Item 6 „Halluzinationen und Wahnvorstellungen“ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Item 7 „Gedrückte Stimmung“ | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ |
| Item 8 „Andere Symptome“ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ |
| Item 9 „Probleme mit Beziehungen“ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ |
| Item 10 „Probleme mit alltäglichen Aktivitäten“ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ |
| Item 11 „Probleme durch die Wohnbedingungen“ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ |
| Item 12 „Probleme im Beruf und in der Freizeit“ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ |
| **Gesamtscore der HoNOS-D** | ✓ | ✓ | k.H. | ✓ | ✓ | ✓ |

**Anmerkungen:** ✓ = Hypothese bestätigt; ✗ = Hypothese verworfen; k.H. = keine Hypothese formuliert; Hypothese 1a „Praktikabilität“; Hypothese 1b „Verteilungseigenschaften“; Hypothese 1c „Interkorrelationen“; Hypothese 1d „Interraterreliabilität“; Hypothese 1e „Test-Retest-Reliabilität“

Ergänzend zu den ausführlichen Analysen zur Praktikabilität und Reliabilität der HoNOS-D sollen nun im Folgenden kurz die Ergebnisse der Analysen zur generellen Validität der HoNOS-D (Harfst et al., 2004a) vorgestellt werden. Da sie nicht zentraler Bestandteil der vorliegenden Arbeit sind, sollen sie an dieser Stelle nur kurz referiert werden.

Harfst et al. (2004a) überprüften die Validität der HoNOS-D anhand der vorliegenden Untersuchungsstichprobe von N = 1918 Patienten. Im Einzelnen wurden die faktorielle Validität der HoNOS-D, ihre konkurrente und diskriminante Validität, die Änderungssensitivität und die prognostische Validität untersucht. Anhand der vorliegenden Untersuchungsstichprobe konnte die von den Testautoren postulierte 4-Faktoren-Lösung der HoNOS-D in einer explo-

rativen Faktorenanalyse nicht repliziert werden. Die Autoren konstatieren, dass dies vor dem Hintergrund der Annahme der weitestgehenden Unabhängigkeit der einzelnen Items untereinander ein erwartungsgemäßes Ergebnis darstellt. Weiterhin konnten Harfst et al. (2004a) erste Hinweise zur konkurrenten Validität der HoNOS-D aufzeigen (z.B. HoNOS-D Gesamtscore[19] und SF-8 Psychische Summenskala $r = .523$). Es ergaben sich auch Korrelationen mit mindestens kleiner Effektstärke zu verschiedenen Skalen anderer selbst eingeschätzter Instrumente (GSI der SCL-14, Subskalen der SF-8-S). Des Weiteren fanden Harfst et al. (2004a), dass der Gesamtscore der HoNOS-D geeignet ist, die Behandlungsdauer von Patienten in Multiplen Regressionsanalysen vorherzusagen ($ß = .131$, $p < .001$).

## 10.3 Die spezifische Eignung der HoNOS-D (Hypothese 2)

### *10.3.1 Die Bedeutung der HoNOS-D zur Fallgruppenbildung (Hypothese 2a)*

Die Überprüfung der Hypothese 2a „Prädiktor HoNOS-D“ zur Bedeutung der HoNOS-D innerhalb der Fallgruppenbildung erfolgte mittels Regression-Tree-Verfahren. Die Anwendungsvoraussetzungen zur Verteilung der Prädiktoren und der Kriteriumsvariablen wurden bereits im Methodenteil (Kap. 9) bzw. für den Gesamtscore der HoNOS-D im Ergebnisteil (Kap. 10, Abschnitt 10.1) vorliegender Arbeit vorgestellt und geprüft.

Abbildung 44 zeigt den mittels Regression-Tree-Verfahren gebildeten Entscheidungs- bzw. **Fallgruppenbaum zur Vorhersage der Behandlungsdauer** von Patienten mit psychischen Störungen in stationärer Rehabilitationsbehandlung. Von den 37 Prädiktoren (vergleiche Tabelle 17 in Abschnitt 9.5.1.1 und Tabelle 19 in Abschnitt 9.5.2.1), die in die Analyse eingingen, trugen 12 Prädiktoren zur Bildung ressourcenhomogener Gruppen bei. Auf oberster Trennungs- bzw. Baumebene erwies sich die Variable „Erstdiagnose bei Aufnahme“ als bedeutsamster Prädiktor. Auf der zweiten Trennungsebene zeigten sich die Merkmale „zusätzli-

[19] In der Analyse von Harfst et al. (2004a) zur Validität der HoNOS-D wurde der Gesamtscore über die ersten 10 Items des Instrumentes gebildet, wobei mindestens 8 Items beantwortet sein mussten.

che Einschätzung des Schweregrades der Erstdiagnose“, „Alter“ und „Chronifizierung der Haupterkrankung“ zur Vorhersage der Behandlungsdauer als relevant. Diesen Trennungsmerkmalen folgen auf der **dritten Ebene** die Variablen „Belastung in der gesundheitsbezogenen Lebensqualität (Fremdeinschätzung)“, **„Schweregrad (HoNOS-D Gesamtscore)“**, „Belastung durch interpersonelle Probleme“ und „Vorhandensein von Komorbidität“. Auf der vierten Baumebene erweisen sich die Prädiktoren „psychotherapeutische Vorerfahrung“ der Patienten, „aktuelle Partnersituation“, „gesundheitsbezogene Lebensqualität (Selbsteinschätzung)“, „Komorbidität“ und „BMI“ der Patienten als bedeutsam. Somit stellt die Schweregradausprägung im HoNOS-D erwartungsgemäß einen Prädiktor innerhalb der Fallgruppenbildung von Patienten mit psychischen Störungen in stationärer Rehabilitationsbehandlung dar.

Weiterhin zeigt sich, dass insgesamt **21 ressourcenhomogene Fallgruppen** gebildet wurden, die eine **Gesamtvarianz von 34.4%** aufklären. Die Patienten mit den durchschnittlich längsten Behandlungsdauern (Abbildung 44, dunkelblau hinterlegt bzw. dunkelgrau, M = 81 Tage, SD = 14) sind dadurch gekennzeichnet, dass sie eine Essstörung als Erstdiagnose aufweisen, die von den Behandlern als besonders schwer beeinträchtigend eingeschätzt wurde, und dass sie sowohl stärker in der fremd eingeschätzten gesundheitsbezogenen Lebensqualität belastet sind als auch über ambulante psychotherapeutische Vorerfahrung (mindestens 6 Termine) verfügen. Dagegen sind die Patienten mit den durchschnittlich kürzesten Behandlungsdauern (Abbildung 44, hellblau hinterlegt bzw. hellgrau, M = 36 Tage, SD = 10) dadurch gekennzeichnet, dass sie eine „Neurotische-, Belastungs- oder somatoforme Störung“ als Erstdiagnose aufweisen, höchstens 5 Jahre erkrankt sind, geringer interpersonell belastet sind und einen verhältnismäßig niedrigen BMI haben (vgl. Abbildung 44). Bei den beiden Gruppen, die hinsichtlich des Ressourcenverbrauches als Endpunkte zu bezeichnen sind, spielt die Schweregradbelastung im HoNOS-D keine Rolle.

Bezüglich der Stellung der **„Gesamtscore HoNOS-D“** als Prädiktor im Fallgruppenbaum zeigt sich, dass die Einschätzung des Schweregrades im HoNOS-D ebenso wie die „Psychische Summenskala des SF-8-F“ auf der dritten Trennungsebene des Baumes von Bedeutung ist. Dieser Befund soll in Abschnitt 10.3.2 mittels einer Linearen Regressionsanalyse differenzierter untersucht werden.

Konkret ergibt sich für den Prädiktor „Gesamtscore der HoNOS-D“, dass bei Patienten mit affektiven Störungen, die älter als 29 Jahre sind, der Schweregrad zur Unterscheidung hinsichtlich des Ressourcenverbrauchs beiträgt. Dabei bildet der Prädiktor „Gesamtscore der HoNOS-D“ drei weitere Gruppen, die bezüglich der durchschnittlichen Behandlungsdauer variieren. Je nach Ausprägung im Schweregrad der HoNOS-D sind dann auf der vierten Trennungsebene des Baumes wiederum drei unterschiedliche Prädiktoren von Bedeutung: „Partnersituation“, „Ausprägung in der gesundheitsbezogenen Lebensqualität (SF-8-S)“ und „Komorbidität“. In Kombination mit diesen drei Merkmalen trägt der „Gesamtwert im HoNOS-D“ zu einer Trennung von insgesamt **6 ressourcenhomogenen Fallgruppen** bei, deren mittlere Behandlungsdauer in einem Bereich zwischen M = 38 und M = 48 Tagen liegt. Betrachtet man die drei Gruppen, in die der Prädiktor „Gesamtscore der HoNOS-D“ trennt, so lässt sich zeigen, dass je höher belastet die Patienten im Schweregrad eingeschätzt werden ($> 1.58$) auch tendenziell längere Behandlungsdauern auftreten ($M_{BD} = 46$ Tage, $SD_{BD} = 11$ Tage; vgl. Abbildung 44, Knoten 16, Knoten 17 und Knoten 18).

Bevor die Ergebnisse des Regression-Tree-Verfahrens im Hinblick auf die zu prüfende Hypothese 2a bewertet werden, sollen sie hinsichtlich ihrer Stabilität überprüft werden. Dazu soll anhand von zufällig gezogenen Stichproben zum einen die extrahierte Baumstruktur (Abschnitt 10.3.1.1) und zum anderen die Behandlungsdauerrangreihung der Fallgruppen (Abschnitt 10.3.1.2) überprüft werden.

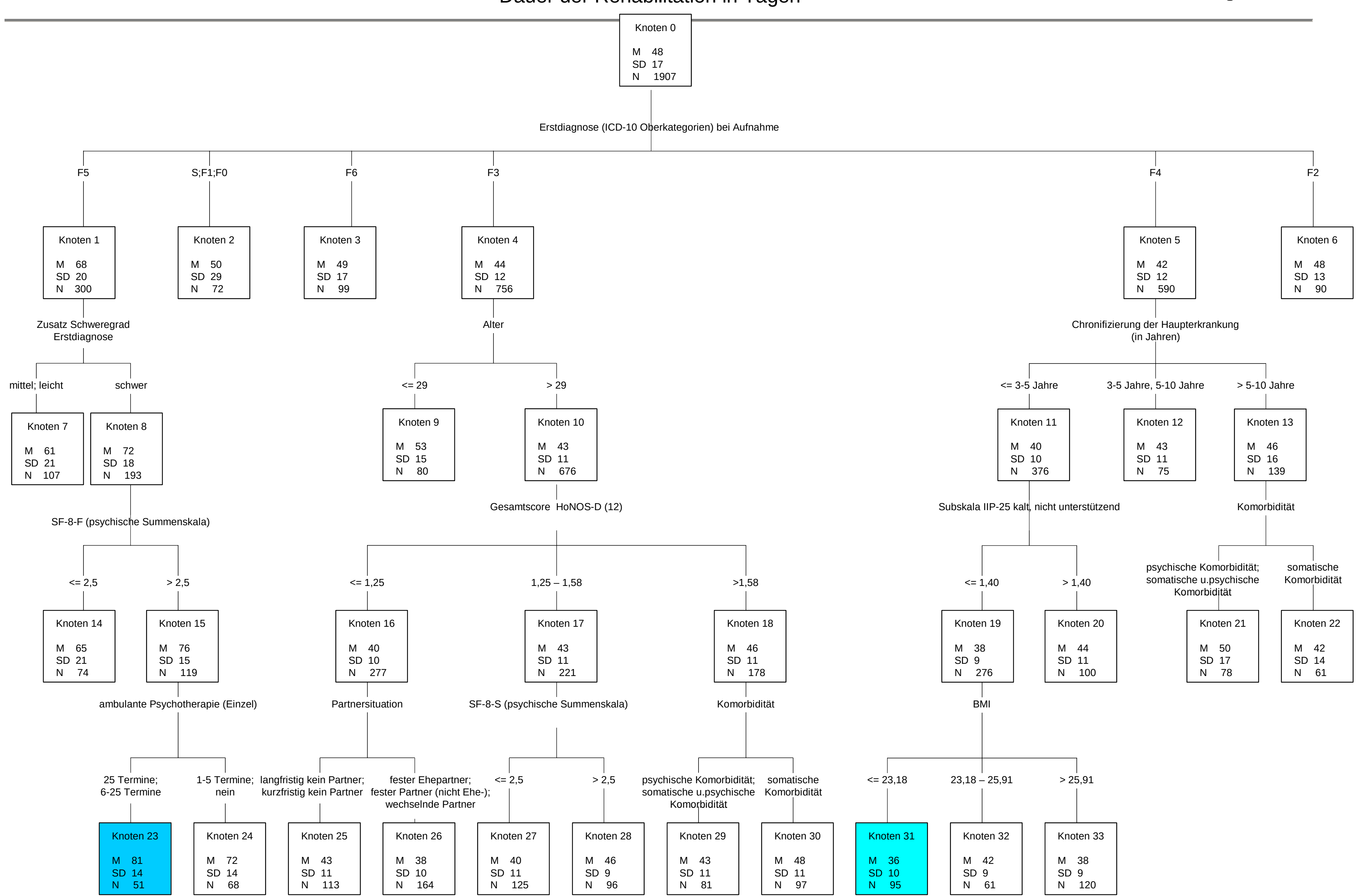



Abbildung 44: Entscheidungsbaum der Kriteriumsvariable „Behandlungsdauer“. Es wurden 21 Fallgruppen mit einer Gesamtvarianzaufklärung von 34,4% gebildet. Die farbig unterlegten Gruppen sind jeweils die Gruppen mit den längsten bzw. kürzesten Behandlungsdauern.

#### 10.3.1.1 Überprüfung der Struktur der Fallgruppen mittels Regression-Tree-Verfahren

Wie bereits im Abschnitt 9.6.2.1 ausgeführt, stellt ein in der Literatur relativ häufig diskutiertes Problem die Variabilität der Baumstrukturen dar. Deswegen wird in der Literatur, wie bereits erwähnt, zur Überprüfung der Stabilität der mittels Regression-Tree-Verfahren extrahierten Merkmale die Methode des „Bootstrapping“ (Ziehen mit Zurücklegen) empfohlen (Schweizer & Vach, 1999). Dabei werden Informationen **einer** empirisch untersuchten Stichprobe verwendet, mit der Zielsetzung, Aussagen zur Variabilität von Stichprobenkennwerten zu gewinnen (Bortz, 1995). Das Bootstrap-Verfahren ist im SPSS Modul Answer Tree 3.1 implementiert (SPSS, 2004) und wird dort über die Ausgabe eines entsprechenden Kennwertes realisiert. Da hier jedoch, über diesen statistischen Kennwert hinausgehend, eine weiterführende Überprüfung der Stabilität der gefundenen Baumstruktur bei Stichprobeninvarianz vorgenommen werden soll, wurde ergänzend ebenfalls ein Bootstrap-Verfahren angewandt (Bortz, 1995).

Die Überprüfung der Stabilität der Baumstruktur erfolgte anhand von zehn zufällig gezogenen Stichproben, für die jeweils 97% der Patienten der Gesamtstichprobe nach dem Zufallsprinzip ausgewählt wurden. Zu jeder Zufallsstichprobe wurde ein Regression-Tree-Verfahren 37 Prädiktoren berechnet, die vormals in die Analyse eingingen (vgl. hierzu auch Tabelle 17 in Abschnitt 9.5.1.1 und Tabelle 19 in Abschnitt 9.5.2.1).

Zur besseren Veranschaulichung und Interpretation der Ergebnisse wurde der Gesamtbaum in einzelne Segmente unterteilt (vgl. Abbildung 45), die sowohl durch die 6 Äste des Baumes (von links nach rechts) als auch durch die vier Ebenen gebildet werden (vgl. Abbildung 44). Anschließend wurde überprüft, in wie weit sich die Baumstrukturen innerhalb der einzelnen Segmente in den zehn Zufallsstichproben als stabil erweisen.

Zur Bewertung der Ergebnisse wurden folgende Kriterien zugrunde gelegt: Eine 100%-ige Übereinstimmung zwischen Hauptbaum und den 10 Bäumen der Zufallsstichproben gilt als sehr gutes Ergebnis; 90 bis 80% Übereinstimmung sind als gut, 70 bis 60% Übereinstimmung als befriedigend und unter 50% Übereinstimmung als moderates Ergebnis zu werten. Aufgrund des rekursiven Vorgehens des Exhaustive-CHAID Algorithmus ist auf den unteren Ebenen der Bäume mit höherer Wahrscheinlichkeit eine niedrigere prozentuale Übereinstim-

mungsrate zu erwarten. Da in den unteren Baumstrukturbereichen vermutlich eine geringere Distanz im Anteil der erklärten Varianz auftritt.

Abbildung 45 illustriert die in 24 Segmente gegliederte Struktur des Hauptbaumes. Tabelle 30 enthält die Übereinstimmungsraten zwischen dem Hauptbaum und den 10 in den Zufallsstichproben gebildeten Entscheidungsbäumen. Zur besseren Nachvollziehbarkeit der berichteten Übereinstimmungsquoten befinden sich die 10 einzelnen Entscheidungsbäume im Anhang 16.5.2.

Wie aus Tabelle 30 und Abbildung 45 ersichtlich, zeigt sich für Ebene 1 (= Ebene des Prädiktors „Erstdiagnose"), dass sich die Erstdiagnosen „F5" („Verhaltensauffälligkeiten mit körperlichen Störungen oder Faktoren") und „F4" („Neurotische-, Belastungs- und somatoforme Störungen") in allen zehn Entscheidungsbäumen der Zufallsstichproben wieder finden und somit zu 100% als stabil erweisen. Die Erstdiagnose „F3" („Affektive Störungen") und „F6" („Schizophrenie, schizotype und wahnhafte Störungen") wird in immerhin 8 von 10 Entscheidungsbäumen extrahiert, wohingegen die Erstdiagnosegruppe „S; F1; F0" („Somatische Komorbidität; „Psychische und Verhaltensstörungen durch psychotrope Substanzen" und „Organische, einschließlich symptomatischer psychischer Störungen") bzw. „F6" („Persönlichkeits- und Verhaltensstörungen") sich nur bedingt als stabil erweisen (jeweils zu 60%).

Auf Ebene 2, die die Prädiktoren „Schweregradbeurteilung der Erstdiagnose", „Alter" und „Chronifizierung der Haupterkrankung" umfasst, zeigen sich gute Ergebnisse mit Übereinstimmungsraten zwischen 80 und 90%, d.h. diese Merkmale wurden in 8 bzw. 9 der 10 Entscheidungsbäume (Zufallsstichproben) identisch extrahiert (siehe Tabelle 30).

Ebene 3 mit den Prädiktoren „Ausprägung im SF-8-F", **„Gesamtscore im HoNOS-D"**, „Ausprägung im IIP-25" und „Komorbidität" werden in den zehn Entscheidungsbäumen der Zufallsstichproben zwischen 60% und 80% der Merkmalskombinationen repliziert. Der „Gesamtscore der HoNOS-D" erscheint in 6 der 10 Entscheidungsbäume als signifikanter Prädiktor des Ressourcenverbrauchs (Ast 4 – Ebene 3, vgl. Abbildung 45 und Tabelle 30).

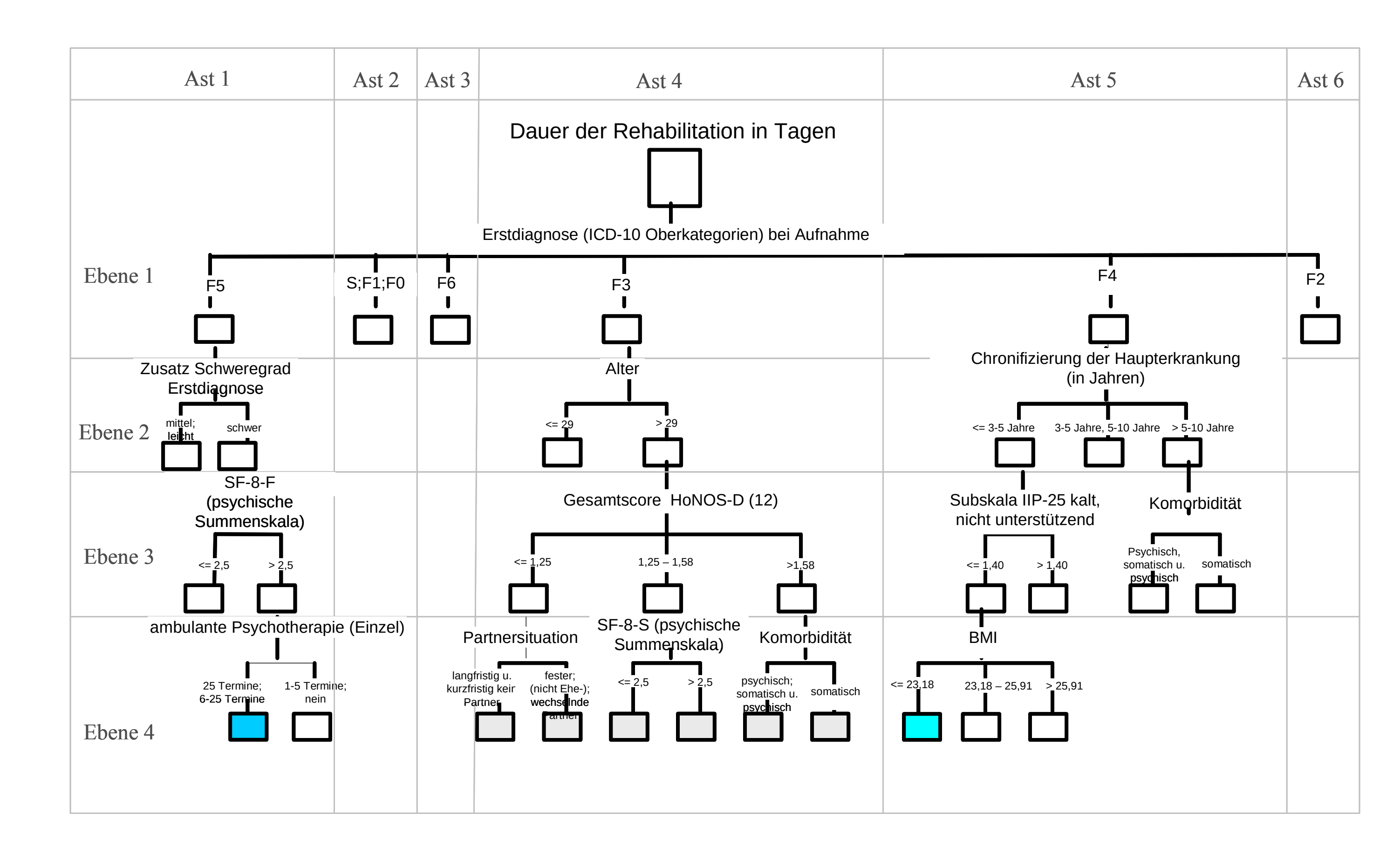



Abbildung 45: Der in 24 Segemente untergliederte Hauptbaum (vgl. hierzu auch Abbildung 44 im vorigen Abschnitt).

Tabelle 30: Prozentualer Anteil der Replikation der Original-Baumstruktur in zehn 97-prozentigen Zufallsbäumen in den verschiedenen Ästen und auf vier Trennungsebenen

| | **Ast 1** | **Ast 2** | **Ast 3** | **Ast 4** | **Ast 5** | **Ast 6** |
|---|---|---|---|---|---|---|
| **Ebene 1** | 100% | 60% | 60% | 80% | 100% | 80% |
| **Ebene 2** | 90% | | | 80% | 90% | |
| **Ebene 3** | 80% | | | **60%** | 80% | |
| **Ebene 4** | 60% | | | 10% | 0% | |

Auf der vierten Ebene, der untersten Trennungsebene, werden wie erwartet überwiegend nur noch moderate bzw. unbefriedigende Übereinstimmungsraten erzielt (vgl. Tabelle 30). Gleichzeitig spricht dies zunächst aber auch für eine unbefriedigende Stabilität für die 11 der 21 Fallgruppen, die sich auf dieser Ebene befinden. Um diese mögliche Einschränkung näher

zu untersuchen, wurde in einer zweiten Analyse eine weitere Überprüfung der Rangfolge hinsichtlich der Behandlungsdauer der extrahierten Fallgruppen vorgenommen (vgl. Abschnitt 10.3.1.2).

Ein weiteres Gütekriterium der Entscheidungsbäume ist die erzielte Varianzaufklärung. Diese bewegte sich für die zehn Bäume der Zufallsstichproben in einem Bereich zwischen 33.83% und 37.11%, in dem auch die extrahierte Varianzaufklärung des Gesamtbaumes mit 34.4% fällt, was als befriedigendes Resultat zu bewerten ist.

In den Entscheidungsbäumen zu den zehn zufällig gezogenen Stichproben ergaben sich auch neu extrahierte Prädiktoren zur Vorhersage der Behandlungsdauer, die der Vollständigkeit halber an dieser Stelle aufgeführt werden sollen. Es handelte sich hierbei um die Merkmale „höchster beruflicher Abschluss“, „aktuelle berufliche Situation“, „gegenwärtige berufliche Position“, „Krankschreibung zum Zeitpunkt der Aufnahme“, „Motivation zur Rehabilitationsbehandlung“, „GSI der SCL-14“, „Einnahme von Antidepressiva“ und „Gesamtwert in der ADS-K“ (siehe Anhang 16.5.2).

**Zusammenfassend** lässt sich zur Stabilität der Baumstruktur aussagen, dass sich anhand von zufällig gebildeten Stichproben bereits auf der obersten Ebene Werte zwischen 60% bis 100% aufzeigen lassen. Problematisch erscheint weiterhin der Umstand, dass sich diejenigen Prädiktoren im Baummodell als instabiler erwiesen, die später bzw. erst auf den unteren Ebenen als statistisch signifikant extrahiert werden. Dies könnte unter Umständen Auswirkungen auf die Stabilität der einzelnen Fallgruppen haben, da sich diese zumeist erst auf der letzten Trennungsebene konstituieren.

##### 10.3.1.2 Überprüfung der Fallgruppen des Fallgruppenbaumes mittels Linearer Regressionsanalyse

Wie oben beschrieben, zeigte sich in der vorausgegangenen Analyse, dass sich 11 der extrahierten Fallgruppen in den zehn Entscheidungsbäumen der Zufallsstichproben als relativ instabil erwiesen haben. Deswegen soll nun eine Überprüfung der Rangfolge der Fallgruppen hinsichtlich ihrer Stichprobeninvarianz, gemessen am Indikator „Anteil aufgeklärter Varianz“ überprüft werden. Die empirischen Analysen folgen hierbei der Fragestellung: Weisen die

Fallgruppen in den Zufallsstichproben (proportional) vergleichbare Behandlungsdauern wie im Originalbaum auf (vgl. Abbildung 44)? Dabei interessierte entsprechend der Hypothese 2a insbesondere die Überprüfung der 6 HoNOS-D bezogenen ressourcenhomogenen Fallgruppen (vgl. Abbildung 44).

In Anlehnung an das Vorgehen von Lemon et al. (2003), die eine logistische Regressionsanalyse (Kriteriumsvariable = Nominalskalenniveau) zur Überprüfung der Stabilität der mittels Regression-Tree-Verfahren gebildeten Gruppen und der extrahierten Varianzaufklärung berechneten, wird in der vorliegenden Arbeit eine Lineare Regressionsanalyse (Kriteriumsvariable „Behandlungsdauer" = Intervallskalenniveau) zur Überprüfung der Rangfolge der Fallgruppen hinsichtlich ihrer durchschnittlichen Behandlungsdauer anhand von *fünf* zufällig gezogenen Stichproben eingesetzt. Die Lineare Regressionsanalyse soll verwendet werden, um über die Rangfolge der unstandardisierten b-Gewichte, welche die Äquivalente zur durchschnittlichen Behandlungsdauer der Fallgruppen in den Zufallsstichproben darstellen[20], eine Aussage zur Stabilität der Fallgruppenrangreihung und - über das $R^2_{adj}$ als Gütemaß der Multiplen Regressionsanalyse - eine Aussage zur Stabilität der Varianzaufklärung treffen zu können (Backhaus et al., 2000).

Um die Darstellbarkeit der Ergebnisse und ihre Interpretierbarkeit besser gewährleisten zu können, wurde die Anzahl der Zufallsstichproben auf fünf festgelegt. Zudem wurde das Kriterium der Zufallsstichprobenbildung verschärft und jeweils nur noch 70% der Patienten aus der Gesamtstichprobe gezogen. Diese Strenge wurde vorgenommen, da es sich bei diesen Berechnungen um kostenrelevante Parameter handelt und die Ergebnisse hinsichtlich der Generalisierbarkeit der extrahierten Fallgruppen bewertet werden sollen.

An dieser Stelle soll ausdrücklich betont werden, dass die hier gewählte Verwendung der Linearen Regressionsanalyse *vom sonst üblichen Verständnis* sowohl in methodischen Voraussetzungen als auch in der Interpretation der Ergebnisse *abweicht*. Der Unterschied besteht

[20] Die durchschnittliche Behandlungsdauer der Fallgruppen in den Zufallsstichproben entspricht der Summe aus unstandardisiertem b-Gewicht und Konstante der Linearen Regressionsanalyse.

zum einen darin, dass die Merkmalskonstellationen der Fallgruppen, die als Prädiktoren in die Lineare Regressionsanalyse eingehen werden, bereits durch die Ergebnisse des Regression-Tree-Verfahrens vorgegeben sind. Zum anderen ist es nicht Ziel dieser Analyse, die Bedeutsamkeit einzelner Prädiktoren zu interpretieren, d.h. die gewonnenen b-Gewichte haben keine weitere interpretatorische Bedeutung, sondern repräsentieren jeweils die durchschnittliche Behandlungsdauer in den einzelnen Fallgruppen der Zufallsstichproben.

Konkret beinhaltete das methodische Vorgehen, dass die Merkmalskonstellationen, so wie sie durch das Ergebnis des Regression-Tree-Verfahrens vorgegeben waren, als dichotome Variablen in die Lineare Regressionsanalyse eingespeist wurden. Alle diejenigen Patienten wurden mit einer 1 kodiert, für die eine Merkmalskonstellation zutraf, alle übrigen Patienten erhielten eine 0.

*Ein Beispiel soll das Vorgehen illustrieren:*

Die Patienten der Fallgruppe 9 (oder des Knotens 9, vgl. hierzu auch Abbildung 44) sind dadurch gekennzeichnet, dass sie als Erstdiagnose eine „Affektive Störung“ aufweisen und jünger als 29 Jahre sind (n = 80). Alle diejenigen Patienten (n = 80), die nun als Erstdiagnose eine „Affektive Störung“ aufwiesen und jünger als 29 Jahre alt waren, bekamen in der Variable „Fallgruppe 9“ eine **1** zugewiesen, alle übrigen Patienten erhielten eine 0 (n = 1827).

Nach Abschluss der Dichotomisierung aller Merkmalskonstellationen waren alle Patienten den 21 Gruppen, so wie sie im Fallgruppenbaum extrahiert wurden (vgl. hierzu auch Abbildung 44), zugeordnet worden. Anschließend wurden fünf Zufallsstichproben gezogen und die 21 dummycodierten Variablen in die Lineare Regressionsanalyse eingegeben.

Die Ergebnisse der empirischen Analysen zu den insgesamt fünf Zufallsstichproben sind in den Tabelle 31 bis Tabelle 35 dargestellt. Zur Verbesserung der Interpretierbarkeit der Ergebnisse wurden die Fallgruppen des Hauptbaumes absteigend in Bezug auf ihre durchschnittliche Behandlungsdauer durchnummeriert. Gruppe 1 enthält nun also die Patienten mit den durchschnittlich längsten Behandlungsdauern. Ebenso wurden die 21 Variablen der Zufallsstichproben auf der Basis ihrer unstandardisierten b-Gewichte in eine absteigende Reihenfolge gebracht. Diese beiden Rangreihungen wurden dann zum direkten Vergleich gegenübergestellt (vgl. hierzu Tabelle 31 bis Tabelle 35). Eine schraffierte Markierung sollte zusätzlich kenntlich machen, wann die Rangreihung nach der Behandlungsdauer des Hauptbaumes nicht mehr mit der Rangreihung der Zufallsstichproben übereinstimmte.

Es soll nun eine zusammenfassende Ergebnisdarstellung über alle fünf Zufallsstichproben erfolgen. Die Stichprobenumfänge der gezogenen Zufallsstichproben variierten in einem Bereich von N = 1327 bis N = 1355. Innerhalb der fünf Stichproben erwiesen sich im Mittel ca. 14 Fallgruppen hinsichtlich ihrer **Behandlungsdauerrangreihung** im Vergleich zur ursprünglichen Reihenfolge als identisch. In der ersten Zufallsstichprobe erwiesen sich 17 Fallgruppen als identisch, dagegen in der dritten Stichprobe nur 11 Gruppen. In 4 der 5 Zufallsstichproben ergab sich für jeweils 1 der 21 Fallgruppen eine starke Abweichung von der ursprünglichen Behandlungsdauerrangreihung. Für die übrigen Fallgruppen in den Zufallsstichproben zeigten sich leichte bzw. moderate Abweichungen.

Von den **6 ressourcenhomogenen Fallgruppen des Hauptbaumes**, die u.a. durch den Prädiktor „Gesamtscore der HoNOS-D" gebildet wurden, erwiesen sich 4 Gruppen (in 4 Analysen) bzw. 5 Gruppen (in 1 Analyse) in den fünf Zufallsstichproben als identisch zur ursprünglichen Behandlungsdauerreihung. In zwei Zufallsstichproben ergab sich für jeweils 1 Gruppe eine starke Abweichung (siehe Tabelle 33 und Tabelle 34). Alle anderen HoNOS-D bezogenen Fallgruppen wiesen in den fünf zufällig gezogenen Stichproben eher leichte Abweichungen im Vergleich zur ursprünglichen Behandlungsdauerrangreihung auf.

Hinsichtlich des untersuchten Kennwertes der Varianzaufklärung lässt sich sagen, dass dieser für die fünf untersuchten zufälligen Stichproben in einem Bereich zwischen 30.8% und 34.0% ($R^2_{adj}$) variierte (siehe Tabelle 31 bis Tabelle 35) und damit nah am Varianzaufklärungsbereich des Hauptbaumes (34.4%) liegt.

**Zusammenfassend** lässt sich für die Übereinstimmung der Behandlungsdauerrangreihung zwischen Fallgruppen des Hauptbaumes und den Fallgruppen der Zufallsstichproben aussagen, dass sich im Mittel ca. 14 der insgesamt 21 Fallgruppen in zufällig gebildeten Stichproben hinsichtlich ihrer Rangreihung als identisch erweisen und nur bei jeweils 1 Fallgruppe starke Abweichungen auftraten. Von den 6 HoNOS-D bezogenen Fallgruppen erweisen sich in den fünf Zufallsstichproben 4 bzw. 5 Fallgruppen hinsichtlich ihrer Behandlungsdauerreihung als identisch.

Tabelle 31: Überprüfung der Stabilität der Original-Fallgruppen (Knoten) in der ersten Zufallsstichprobe (70% der Gesamtstichprobe) in absteigender Rangreihung nach Behandlungsdauer und undstandardisierter b-Gewichte (N = 1.335)

| Knoten | Rang im Hauptbaum | Rangreihe BD im Hauptbaum | Rang im Hauptbaum | Unstandardisierte b-Gewichte | Rangreihe BD nach Zufallsstichprobe |
|---|---|---|---|---|---|
| 23 | 1 | 80.98 | 1 | 43.59 | 81.03 |
| 24 | 2 | 71.54 | 2 | 34.65 | 72.09 |
| 14 | 3 | 64.99 | 3 | 27.20 | 64.64 |
| 7 | 4 | 61.38 | 4 | 22.65 | 60.09 |
| 9 | 5 | 52.69 | 5 | 17.49 | 54.92 |
| 2 | 6 | 50.13 | 6 | 14.25 | 51.69 |
| 21 | 7 | 49.95 | 7 | 12.78 | 50.22 |
| 3 | 8 | 48.59 | 8 | 11.81 | 49.24 |
| **30** | 9 | 48.30 | 9 | 10.12 | 47.55 |
| 6 | 10 | 47.72 | 10 | 9.04 | 46.48 |
| **28** | 11 | 45.84 | 11 | 8.77 | 46.21 |
| 20 | 12 | 43.76 | 15 | 6.54 | 43.98 |
| **29** | 13 | 43.48 | 13 | 6.21 | 43.65 |
| **25** | 14 | 43.07 | 12 | 5.62 | 43.06 |
| 12 | 15 | 42.65 | 16 | 5.44 | 42.88 |
| 22 | 16 | 41.67 | 14 | 5.08 | 42.52 |
| 32 | 17 | 41.64 | 17 | 4.18 | 41.62 |
| **27** | 18 | 40.12 | 18 | 3.30 | 40.73 |
| **26** | 19 | 38.37 | Konstante | .000 | 37.44 |
| 33 | 20 | 37.93 | 20 | 1.54 | 38.97 |
| 31 | 21 | 35.73 | 21 | -1.69 | 35.75 |
| | | | | | |
| R | .577 | | | | |
| R² | **.333** | | | | |
| R²adj | .323 | | | | |
| F | 32.81 | | | | |
| df | 20/1314 | | | | |
| p | < .001 | | | | |

Anmerkungen: weiß = identisch zur ursprünglichen Behandlungsdauerreihung; hellgrau = leicht abweichend; dunkelgrau = stärker abweichend; schwarz = stark abweichend; fett markierte Knoten = HoNOS-Fallgruppen; FG = Fallgruppe; BD = Behandlungsdauer

Tabelle 32: Überprüfung der Stabilität der Original-Fallgruppen (Knoten) in der zweiten Zufallsstichprobe (70% der Gesamtstichprobe) in absteigender Rangreihung nach Behandlungsdauer und unstandardisierter b-Gewichte (N = 1.327)

| **Knoten** | **Rang im Haupt-baum** | **Rangreihe BD im Hauptbaum** | **Rang im Haupt-baum** | **unstandardisierte b-Gewichte** | **Rangreihe BD nach Zufalls-stichprobe** |
|---|---|---|---|---|---|
| 23 | 1 | 80.98 | 1 | 40.15 | 78.35 |
| 24 | 2 | 71.54 | 2 | 31.26 | 69.46 |
| 14 | 3 | 64.99 | 3 | 27.22 | 65.42 |
| 7 | 4 | 61.38 | 4 | 23.10 | 61.30 |
| 9 | 5 | 52.69 | 6 | 14.91 | 53.11 |
| 2 | 6 | 50.13 | 5 | 13.14 | 51.34 |
| 21 | 7 | 49.95 | 7 | 11.37 | 49.57 |
| 3 | 8 | 48.59 | 8 | 10.83 | 49.03 |
| **30** | 9 | 48.30 | 9 | 9.84 | 48.04 |
| 6 | 10 | 47.72 | 10 | 9.44 | 47.64 |
| **28** | 11 | 45.84 | 11 | 7.83 | 46.03 |
| 20 | 12 | 43.76 | 12 | 6.19 | 44.39 |
| **29** | 13 | 43.48 | 14 | 5.27 | 43.47 |
| **25** | 14 | 43.07 | 15 | 4.88 | 43.08 |
| 12 | 15 | 42.65 | 13 | 4.23 | 42.43 |
| 22 | 16 | 41.67 | 17 | 3.23 | 41.43 |
| 32 | 17 | 41.64 | 16 | 2.56 | 40.76 |
| **27** | 18 | 40.12 | 18 | 1.25 | 39.45 |
| **26** | 19 | 38.37 | Konstante | .000 | 38.20 |
| 33 | 20 | 37.93 | 20 | -0.38 | 37.82 |
| 31 | 21 | 35.73 | 21 | -2.33 | 35.87 |
| | | | | | |
| R | .578 | | | | |
| $R^2$ | .334 | | | | |
| $R^2$adj | .324 | | | | |
| F | 32.81 | | | | |
| df | 20/1306 | | | | |
| p | < .001 | | | | |

Anmerkungen: weiß = identisch zur ursprünglichen Behandlungsdauerreihung; hellgrau = leicht abweichend; dunkelgrau = stärker abweichend; schwarz = stark abweichend; fett markierte Knoten = HoNOS-Fallgruppen; FG = Fallgruppe; BD = Behandlungsdauer

Tabelle 33: Überprüfung der Stabilität der Original-Fallgruppen (Knoten) in der dritten Zufallsstichprobe (70% der Gesamtstichprobe) in absteigender Rangreihung nach Behandlungsdauer und unstandardisierter b-Gewichte (N = 1.333)

| Knoten | Rang im Haupt-baum | Rangreihe BD im Hauptbaum | | Rang im Hauptbaum | Unstandardisierte b-Gewichte | Rangreihe BD nach Zufalls-stichprobe |
|---|---|---|---|---|---|---|
| 23 | 1 | 80.98 | | 1 | 42.22 | 80.62 |
| 24 | 2 | 71.54 | | 2 | 33.82 | 72.23 |
| 14 | 3 | 64.99 | | 4 | 24.58 | 62.99 |
| 7 | 4 | 61.38 | | 3 | 24.23 | 62.63 |
| 9 | 5 | 52.69 | | 7 | 12.70 | 51.11 |
| 2 | 6 | 50.13 | | 5 | 12.49 | 50.89 |
| 21 | 7 | 49.95 | | 6 | 10.44 | 48.85 |
| 3 | 8 | 48.59 | | 8 | 10.30 | 48.71 |
| **30** | 9 | 48.30 | | 9 | 9.87 | 48.28 |
| 6 | 10 | 47.72 | | 10 | 9.21 | 47.62 |
| **28** | 11 | 45.84 | | 11 | 7.02 | 45.43 |
| 20 | 12 | 43.76 | | 13 | 6.31 | 44.72 |
| **29** | 13 | 43.48 | | 12 | 5.47 | 43.87 |
| **25** | 14 | 43.07 | | 17 | 4.42 | 42.83 |
| 12 | 15 | 42.65 | | 14 | 4.32 | 42.72 |
| 22 | 16 | 41.67 | | 16 | 3.51 | 41.91 |
| 32 | 17 | 41.64 | | 15 | 3.24 | 41.64 |
| **27** | 18 | 40.12 | | 18 | 2.13 | 40.54 |
| **26** | 19 | 38.37 | | Konstante | .000 | 38.41 |
| 33 | 20 | 37.93 | | 20 | 0.12 | 38.52 |
| 31 | 21 | 35.73 | | 21 | -1.92 | 36.49 |
| | | | | | | |
| R | .591 | | | | | |
| $R^2$ | .350 | | | | | |
| $R^2adj$ | .340 | | | | | |
| F | 35.268 | | | | | |
| df | 20/1312 | | | | | |
| p | < .001 | | | | | |

Anmerkungen: weiß = identisch zur ursprünglichen Behandlungsdauerreihung; hellgrau = leicht abweichend; dunkelgrau = stärker abweichend; schwarz = stark abweichend; fett markierte Knoten = HoNOS-Fallgruppen; FG = Fallgruppe; BD = Behandlungsdauer

Tabelle 34: Überprüfung der Stabilität der Original-Fallgruppen (Knoten) in der vierten Zufallsstichprobe (70% der Gesamtstichprobe) in absteigender Rangreihung nach Behandlungsdauer und unstandardisierter b-Gewichte (N = 1.345)

| Knoten | Rang im Haupt-baum | Rangreihe BD im Hauptbaum | | Rang im Haupt-baum | Unstandardisierte b-Gewichte | Rangreihe BD nach Zufalls-stichprobe |
|---|---|---|---|---|---|---|
| 23 | 1 | 80.98 | | 1 | 40.67 | 79.15 |
| 24 | 2 | 71.54 | | 2 | 32.30 | 70.78 |
| 14 | 3 | 64.99 | | 3 | 25.50 | 63.98 |
| 7 | 4 | 61.38 | | 4 | 21.10 | 59.58 |
| 9 | 5 | 52.69 | | 5 | 13.55 | 52.04 |
| 2 | 6 | 50.13 | | 6 | 12.71 | 51.19 |
| 21 | 7 | 49.95 | | 7 | 12.61 | 51.09 |
| 3 | 8 | 48.59 | | 9 | 10.23 | 48.71 |
| **30** | 9 | 48.30 | | 8 | 9.97 | 48.46 |
| 6 | 10 | 47.72 | | 10 | 9.04 | 47.53 |
| **28** | 11 | 45.84 | | 11 | 7.01 | 45.49 |
| 20 | 12 | 43.76 | | 12 | 5.10 | 43.58 |
| **29** | 13 | 43.48 | | 16 | 5.01 | 43.49 |
| **25** | 14 | 43.07 | | 14 | 4.97 | 43.46 |
| 12 | 15 | 42.65 | | 13 | 4.92 | 43.40 |
| 22 | 16 | 41.67 | | 15 | 4.57 | 43.06 |
| 32 | 17 | 41.64 | | 17 | 2.89 | 41.37 |
| **27** | 18 | 40.12 | | 18 | 2.18 | 40.67 |
| **26** | 19 | 38.37 | | Konstante | .000 | 38.48 |
| 33 | 20 | 37.93 | | 20 | -0.66 | 37.83 |
| 31 | 21 | 35.73 | | 21 | -2.89 | 35.59 |
| | | | | | | |
| R | .573 | | | | | |
| $R^2$ | .329 | | | | | |
| $R^2adj$ | .319 | | | | | |
| F | 35.436 | | | | | |
| df | 20/1324 | | | | | |
| p | < .001 | | | | | |

Anmerkungen: weiß = identisch zur ursprünglichen Behandlungsdauerreihung; hellgrau = leicht abweichend; dunkelgrau = stärker abweichend; schwarz = stark abweichend; fett markierte Knoten = HoNOS-Fallgruppen; FG = Fallgruppe; BD = Behandlungsdauer

Tabelle 35: Überprüfung der Stabilität der Original-Fallgruppen (Knoten) in der fünften Zufallsstichprobe (70% der Gesamtstichprobe) in absteigender Rangreihung nach Behandlungsdauer und unstandardisierter b-Gewichte (N = 1.355)

| Knoten | Rang im Haupt-baum | Rangreihe BD im Hauptbaum | Rang im Haupt-baum | Unstandardisierte b-Gewichte | Rangreihe BD nach Zufalls-stichprobe |
|---|---|---|---|---|---|
| 23 | 1 | 80.98 | 1 | 40.78 | 78.61 |
| 24 | 2 | 71.54 | 2 | 34.65 | 72.48 |
| 14 | 3 | 64.99 | 3 | 26.19 | 64.02 |
| 7 | 4 | 61.38 | 4 | 22.91 | 60.75 |
| 9 | 5 | 52.69 | 5 | 14.83 | 52.66 |
| 2 | 6 | 50.13 | 6 | 14.03 | 51.86 |
| 21 | 7 | 49.95 | 7 | 12.83 | 50.67 |
| 3 | 8 | 48.59 | 9 | 12.18 | 50.02 |
| **30** | 9 | 48.30 | 8 | 11.07 | 48.90 |
| **6** | **10** | **47.72** | **10** | **9.69** | **47.52** |
| **28** | **11** | **45.84** | **11** | **8.40** | **46.23** |
| 20 | 12 | 43.76 | 13 | 6.71 | 44.54 |
| **29** | 13 | 43.48 | 15 | 5.91 | 43.74 |
| **25** | 14 | 43.07 | 14 | 5.07 | 42.90 |
| 12 | 15 | 42.65 | 16 | 4.34 | 42.17 |
| 22 | 16 | 41.67 | 12 | 4.07 | 41.90 |
| 32 | 17 | 41.64 | 17 | 3.69 | 41.52 |
| **27** | 18 | 40.12 | 18 | 3.22 | 41.06 |
| **26** | 19 | 38.37 | Konstante | | 37.83 |
| 33 | 20 | 37.93 | 20 | 0.16 | 37.99 |
| 31 | 21 | 35.73 | 21 | -1.67 | 36.16 |
| | | | | | |
| R | .564 | | | | |
| $R^2$ | .318 | | | | |
| $R^2$adj | .308 | | | | |
| F | 31.090 | | | | |
| df | 20/1334 | | | | |
| p | < .001 | | | | |

Anmerkungen: weiß = identisch zur ursprünglichen Behandlungsdauerreihung; hellgrau = leicht abweichend; dunkelgrau = stärker abweichend; schwarz = stark abweichend; fett markierte Knoten = HoNOS-Fallgruppen; FG = Fallgruppe; BD = Behandlungsdauer

**Zusammenfassung zur Hypothese 2a „Prädiktor HoNOS-D“**

Es wird erwartet, dass der „Gesamtscore der HoNOS-D“ mittels Regression-Tree-Verfahren extrahiert wird und somit einen Prädiktor innerhalb der Fallgruppenbildung darstellt.

Wie die Ergebnisse zeigen, wird der „Gesamtscore der HoNOS-D“ mittels Regression-Tree-Verfahren extrahiert und stellt damit einen Prädiktor innerhalb der Fallgruppenbildung von Patienten mit psychischen Störungen in stationärer Rehabilitationsbehandlung dar. Damit kann die Nullhypothese verworfen werden und die Alternativhypothese als angenommen gelten.

Da die Hypothese 2a bestätigt werden konnte, sollen im Folgenden die Hypothesen 2b und 2c getestet werden.

### *10.3.2 Zur Stellung der HoNOS-D im Vergleich zum SF-8-F innerhalb der Fallgruppenbildung (Hypothese 2b)*

Zur Überprüfung der Hypothese 2b „Bedeutsamkeit des Prädiktors HoNOS-D“ wurden zunächst Korrelationen (Produkt-Moment-Korrelationen) zwischen dem „Gesamtscore der HoNOS-D“ sowie der „Psychischen Summenskala der SF-8-F“ berechnet. Anschließend wurde eine Lineare Regressionsanalyse durchgeführt.

Die Ergebnisse zeigen, dass sowohl der „Gesamtscore der HoNOS-D“ als auch der Gesamtwert der „Psychischen Summenskala der SF-8-F“ signifikant mit kleiner Effektstärke mit der Behandlungsdauer korreliert sind ($r_{BD\text{-}Gesamt\ HoNOS\text{-}D} = .245$, $p \leq .001$, $n = 1897$; $r_{BD\text{-}SF\text{-}8\text{-}F} = .183$, $p \leq .001$, $n = 1893$). Dabei weist der Gesamtscore der HoNOS-D einen deskriptiv höheren Zusammenhang zur Behandlungsdauer auf als der Gesamtscore der SF-8-F. Dieser Unterschied ist nicht signifikant ($z\text{-Wert}_{(Differenz)} = .073$).

Um eine Überprüfung der Bedeutsamkeit des Prädiktors „Gesamtscore HoNOS-D“ im Vergleich zum Prädiktor „Psychische Summenskala der SF-8-F“ vornehmen zu können, gingen alle Prädiktoren, die sich bereits im Regression-Tree-Verfahren als bedeutsam herausgestellt hatten (vgl. Abbildung 44), in eine Multiple Regressionsanalyse ein. Dazu wurde das Standardverfahren „Einschluss“ der Multiplen Regressionsanalyse benutzt (Barth, 1997). Anhand

der standardisierten Regressionskoeffizienten soll anschließend eine Bewertung der Bedeutsamkeit des „Gesamtscores HoNOS-D“ im Vergleich zur „Psychischen Summenskala der SF-8-F“ vorgenommen werden. Die die Anwendungsvoraussetzungen zur Linearen Regressionsanalyse wurden bereits in Abschnitt 9.6.2.2 überprüft.

Die Ergebnisse der Linearen Regressionsanalyse zur Überprüfung der Hypothese 2b „Bedeutsamkeit des Prädiktors HoNOS-D“ sind in Tabelle 36 dargestellt.

Die extrahierte Baumstruktur des Regression-Tree-Verfahrens bestätigend, zeigt sich bei **allgemeiner Betrachtung der Prädiktoren** auch in der Linearen Regressionsanalyse, die Bedeutsamkeit des Prädiktors „Erstdiagnose“ zur Vorhersage der Behandlungsdauer (siehe Tabelle 36). Dabei weist die Mehrzahl der Prädiktoren, die sich auf die Erstdiagnose beziehen, signifikant negative standardisierte Regressionsgewichte auf und ist somit mit längeren Behandlungsdauern assoziiert. Die Prädiktoren „Essstörung“ und „Somatische Erkrankung“ tragen hingegen nicht signifikant zur Vorhersage der Behandlungsdauer bei. Weiterhin zeigt sich, dass diejenigen Prädiktoren, die im Regression-Tree-Verfahren auf den ersten beiden Trennungsebenen als signifikante Merkmale der Behandlungsdauer extrahiert wurden (vgl. Abbildung 44), tendenziell auch in der Multiplen Regressionsanalyse einen höheren Erklärungsanteil an der Behandlungsdauer besitzen als diejenigen Prädiktoren, die im Regression-Tree-Verfahren auf den unteren beiden Trennungsebenen extrahiert wurden. Das lässt sich zum Beispiel für den Prädiktor „Diagnose“ (F3, F4, F6, F2) und den Prädiktor „Alter“ für die oberen Trennungsebenen des Regression-Tree-Verfahrens zeigen. Im Gegenzug besitzen die Prädiktoren der unteren Ebenen des Entscheidungsbaumverfahrens in der Multiplen Regressionsanalyse einen eher geringeren Anteil an der Vorhersage der Behandlungsdauer (z.B. „Komorbidität“ oder „Partnersituation“, siehe Tabelle 36).

Hinsichtlich des direkten **Vergleiches** zwischen der Bedeutung des **„Gesamtscores der HoNOS-D“** und der **„Psychischen Summenskala der SF-8-F“** lässt sich zunächst aussagen, dass beide Merkmale signifikante Prädiktoren zur Vorhersage der Behandlungsdauer darstellen. Darüber hinaus weist der Prädiktor „Psychische Summenskala der SF-8-F“ deskriptiv ein leicht höheres standardisiertes ß-Gewicht auf (ß = .074, p = .002) als der Prädiktor „Gesamtscore der HoNOS-D“ (ß = .067, p = .001). Die beiden standardisierten Koeffizienten der Prädiktoren differieren allerdings nur um einen Betrag von .008 (vgl. Tabelle 36).

Zur Prüfung der Unterschiedlichkeit der beiden nicht-standardisierten Regressionskoeffizienten wurde die kritische Differenz (Amelang & Zielinski, 1997) verwendet. Das methodische Vorgehen wurde bereits in Abschnitt 9.6.2.2 ausführlich beschrieben. Nach den Formeln ergeben sich folgende Werte: $Diff_{krit} = 1.137$ und $Diff_{emp} = 0.148$ ($Standardmessfehler_{SF\text{-}8\text{-}F} = 0.423$; $Standardmessfehler_{HoNOS\text{-}D} = 0.397$). Da $Diff_{emp} < Diff_{krit}$, kann geschlussfolgert werden, dass sich die beiden Regressionskoeffizienten der Prädiktoren „Gesamtscore HoNOS-D" und „Psychische Summenskala der SF-8-F" in dieser Analyse nicht signifikant voneinander unterscheiden.

Ein zusätzlicher Befund besteht in der extrahierten Varianzaufklärung der Multiplen Regressionsanalyse. Es zeigt sich, dass die Varianzaufklärung der Multiplen Regressionsanalyse ($R^2_{adj} = 34.3$) exakt derjenigen Varianzaufklärung entspricht, die durch das Regression-Tree-Verfahren ermittelt werden konnte (Varianzaufklärung von 34.4%).

**Zusammenfassung der Ergebnisse zur Hypothese 2b „Bedeutsamkeit des Prädiktors HoNOS-D"**

**Hypothese 2b:** In der linearen Regressionsanalyse zur Vorhersage der Behandlungsdauer werden höhere signifikante positive Regressionsgewichte für den „Gesamtscore der HoNOS-D" im Vergleich zu einem äquivalenten Fremdeinschätzungsscore „Psychische Summenskala der SF-8-F" erwartet.

Da sich in der Linearen Regressionsanalyse für den „Gesamtscore der HoNOS-D" im Vergleich zur „Psychischen Summenskala der SF-8-F" kein signifikant höheres positives Regressionsgewicht nachweisen lässt, muss die Hypothese 2b hiermit verworfen werden.

Tabelle 36: Standardisierte- und nicht-standardisierte Regressionskoeffizienten, t-Werte, Signifikanzniveaus und Toleranzwerte der Prädiktoren des Regression-Tree-Verfahrens zur Vorhersage der Behandlungsdauer in absteigender Reihenfolge der Beträge der standardisierten β-Gewichte (N = 1.907)

| Prädiktoren | B | Standardisierte ß-Gewichte | t-Wert | p-Wert | Toleranz |
|---|---|---|---|---|---|
| Konstante | 41.207 | | 7.77 | .000 | |
| F3* | -15.05 | -0.423 | -3.32 | 0.001 | 0.02 |
| F4* | -15.64 | -0.415 | -3.43 | 0.001 | 0.02 |
| F6* | -13.19 | -0.168 | -2.79 | 0.005 | 0.09 |
| F2* | -12.41 | -0.151 | -2.62 | 0.009 | 0.10 |
| F5* | 5.70 | 0.119 | 1.24 | 0.21 | 0.04 |
| Partnersituation* (langfristig kein Partner) | 4.20 | 0.100 | 4.67 | .000 | 0.74 |
| Alter | -0.11 | -0.085 | -3.33 | 0.001 | 0.53 |
| S* | -8.73 | -0.078 | -1.74 | 0.082 | 0.17 |
| **SF-8-F (psych. S.)** | **1.308**$^{+}$ | **0.075** | **3.09** | **0.002** | **0.58** |
| Partnersituation* (kurzfristig kein Partner) | 3.78 | 0.069 | 3.30 | 0.001 | 0.78 |
| SF-8-S (psych. S.) | 1.33 | 0.068 | 3.11 | 0.002 | 0.72 |
| **HoNOS-D (Gesamt)** | **1.16**$^{+}$ | **0.067** | **2.93** | **0.003** | **0.66** |
| F1* | -13.90 | -0.066 | -2.32 | 0.021 | 0.43 |
| Partnersituation* (wechselnde Partner) | 2.57 | 0.062 | 2.92 | 0.004 | 0.76 |
| Chronifizierung | 0.72 | 0.059 | 2.98 | 0.003 | 0.88 |
| Ambulante Psychotherapie* (Missing) | 2.86 | 0.056 | 2.88 | 0.004 | 0.90 |
| BMI | 0.12 | 0.053 | 2.62 | 0.009 | 0.83 |
| Komorbidität* (som.) | -1.93 | -0.048 | -2.09 | 0.036 | 0.66 |
| Ambulante Psychotherapie* (mehr als 25 Termine) | 2.28 | 0.042 | 2.03 | 0.042 | 0.80 |
| Partnersituation* (Missing) | 1.47 | 0.030 | 1.53 | 0.127 | 0.91 |
| Schweregradzusatz der Erstdiagnose | 0.81 | 0.027 | 1.31 | 0.189 | 0.80 |
| F0* | -13.41 | -0.025 | -1.22 | 0.222 | 0.82 |
| Komorbidität* (Missing) | 0.76 | 0.020 | 0.99 | 0.320 | 0.88 |
| Komorbidität* (som. und psych.) | -1.20 | -0.020 | -1.00 | 0.315 | 0.85 |
| Ambulante Psychotherapie* (1-5 Termine) | 0.65 | 0.013 | 0.64 | 0.522 | 0.80 |
| Partnersituation* (fester Partner) | 3.04 | 0.06 | 2.83 | 0.005 | 0.77 |
| Partnersituation* (wechselnde Partner) | 0.84 | 0.006 | 0.33 | 0.745 | 0.89 |
| IIP-25 (kalt/nicht-unterstützend) | 0.05 | 0.002 | 0.11 | 0.914 | 0.91 |
| R | .599 | | | | |
| R² | .358 | | | | |
| R²adj | .343 | | | | |
| F | 37.464 | | | | |
| df | 28/1878 | | | | |
| p | < .001 | | | | |

Anmerkungen: * dummycodierte Variablen; $^{+}$ z-standardardisierte Skalenwerte; S = Somatische Komorbidität

### *10.3.3 Die spezifische Eignung der HoNOS-D im Vergleich zur SF-8-F (Psychische Summenskala) (Hypothese 2c)*

Zur Überprüfung der Hypothese 2c „Position im Prädiktoren-Kern-Set“ wurden nun erneut Regression-Tree-Verfahren und Multiple Regressionsanalysen berechnet. Dabei interessierte innerhalb eines „Prädiktoren-Kern-Sets“ (siehe unten) die Extraktion der beiden Merkmale „Gesamtscore HoNOS-D“ und „Psychische Summenskala der SF-8-F“ und deren Beitrag zur erzielten Varianzaufklärung. Zum anderen interessierte (analog zum bisherigen Vorgehen) der Direktvergleich der nicht-standardisierten Regressionskoeffizienten des „Gesamtscores der HoNOS-D“ und der „Psychischen Summenskala der SF-8-F“ als Ergebnis einer Multiplen Regressionsanalyse.

Das **„Prädiktoren-Kern-Set“** wurde unter zwei Prämissen gebildet. Zum ersten wurde es aus den 12 Prädiktoren gebildet, die sich bereits im Regression-Tree-Verfahren als signifikant erwiesen haben (vgl. Abbildung 44) und zum zweiten wurden aus diesen 12 Prädiktoren diejenigen ausgewählt, die am ehesten bereits in der Routineerhebung im Rahmen der Qualitätssicherung von psychosomatischen/psychotherapeutischen Fachkliniken vorhanden sein könnten. Es wurden folgende sechs Merkmale als „Prädiktoren-Kern-Set“ definiert: „Erstdiagnose“, „Alter“, „Partnersituation“, „ambulante psychotherapeutische Vorerfahrung“, „Komorbidität“ und „Chronifizierung der Erkrankung“. Darüber hinaus wurde in einer ersten Analyse der Prädiktor „Gesamtscore der HoNOS-D“ und in einer zweiten Analyse der Prädiktor „Psychische Summenskala der SF-8-F“ in das Regression-Tree-Verfahren eingespeist.

Abbildung 46 und Abbildung 47 zeigen die beiden durch das Regression-Tree-Verfahren extrahierten Baumstrukturen. In der **ersten Analyse**, in der der **„Gesamtscore HoNOS-D“** zusätzlich zum „Prädiktoren-Kern-Set“ eingespeist wurde, wurden 15 Fallgruppen mit einer Gesamtvarianzaufklärung von 31.82% extrahiert (vgl. Abbildung 46). Die **zweite Analyse**, in die das „Prädiktoren-Kern-Sets“ und die **„Psychische Summenskala der SF-8-F“** eingespeist wurden, bildet dagegen 14 Fallgruppen mit einer Gesamtvarianzaufklärung von 31.76% (vgl. Abbildung 47). Damit erweisen sich die beiden variierten Prädiktoren hinsichtlich der Anzahl der extrahierten Fallgruppen und der Varianzaufklärungskennwerte vergleichbar, wobei im Entscheidungsbaum des „Gesamtscores HoNOS-D“ 1 Fallgruppe mehr gebildet, aber eine vergleichbare Varianzaufklärung erzielt wird.

Bezüglich der **Position im Fallgruppenbaum** lässt sich zeigen, dass der Prädiktor **„Gesamtscore der HoNOS-D“** in der zentralen Diagnosegruppe „Affektive Störungen (F3)“ bereits an zweiter Stelle nach dem Prädiktor „Alter“ extrahiert wird. Im Vergleich dazu, erweist sich der Prädiktor **„Psychische Summenskala der SF-8-F“** erst bei dieser Diagnosegruppe nach den Prädiktoren „Alter“ und „Partnersituation“ als relevant. Zusätzlich stellt die „Psychische Summenskala der SF-8-F“ allerdings noch innerhalb der Diagnosegruppe „Essstörung (F5)“ ein bedeutsames Merkmal zur Vorhersage der Behandlungsdauer dar (vgl. Abbildung 46 und Abbildung 47).

In Ergänzung zu den zwei durchgeführten Regression-Tree-Verfahren wurden zwei Multiple Regressionsanalysen gerechnet, um auch hier die Bedeutsamkeit der beiden Prädiktoren („Gesamtscore HoNOS-D“ und „Psychische Summenskala der SF-8-F“) anhand der standardisierten Regressionskoeffizienten zu untersuchen. Da in diesen Analysen, ebenso wie in den vorausgegangenen, die gleiche Kriteriumsvariable und Prädiktorvariablen verwendet wurden, kann an dieser Stelle auf die Ausführungen zur Prüfung der Voraussetzungen der Multiplen Regressionsanalyse in Abschnitt 9.5.1 und 9.5.2 vorliegender Arbeit verwiesen werden. In die Multiplen Regressionsanalysen ging das oben beschriebene „Prädiktoren-Kern-Set“ ein, wiederum zum einen um den „Gesamtscore der HoNOS-D“ und zum anderen um die „Psychische Summenskala der SF-8-F“ ergänzt.

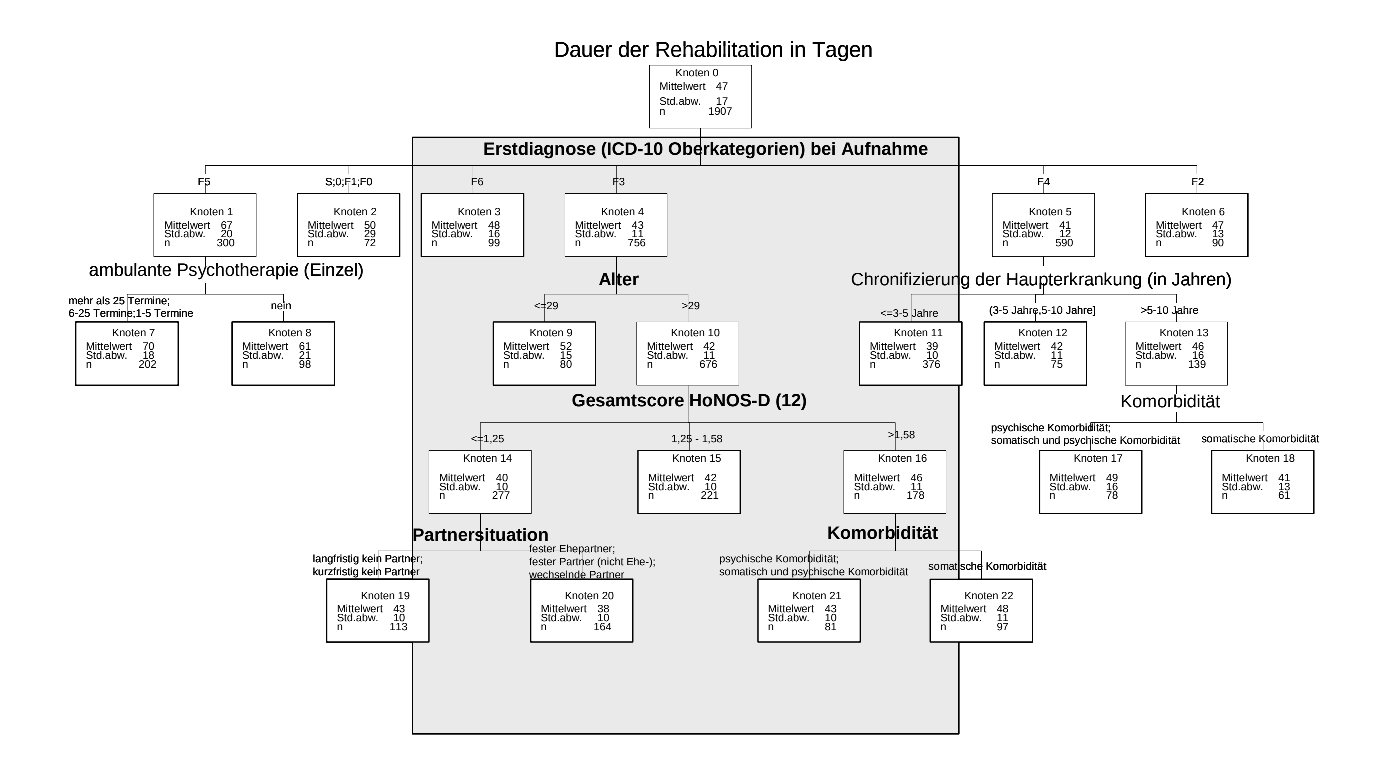



Abbildung 46: Entscheidungsbaum des „Prädiktoren-Kern-Sets“ mit dem zusätzlichen Prädiktor „Gesamtscore der HoNOS-D“. Es wurden 15 Fallgruppen mit einer Gesamtvarianzaufklärung von 31.82% extrahiert.

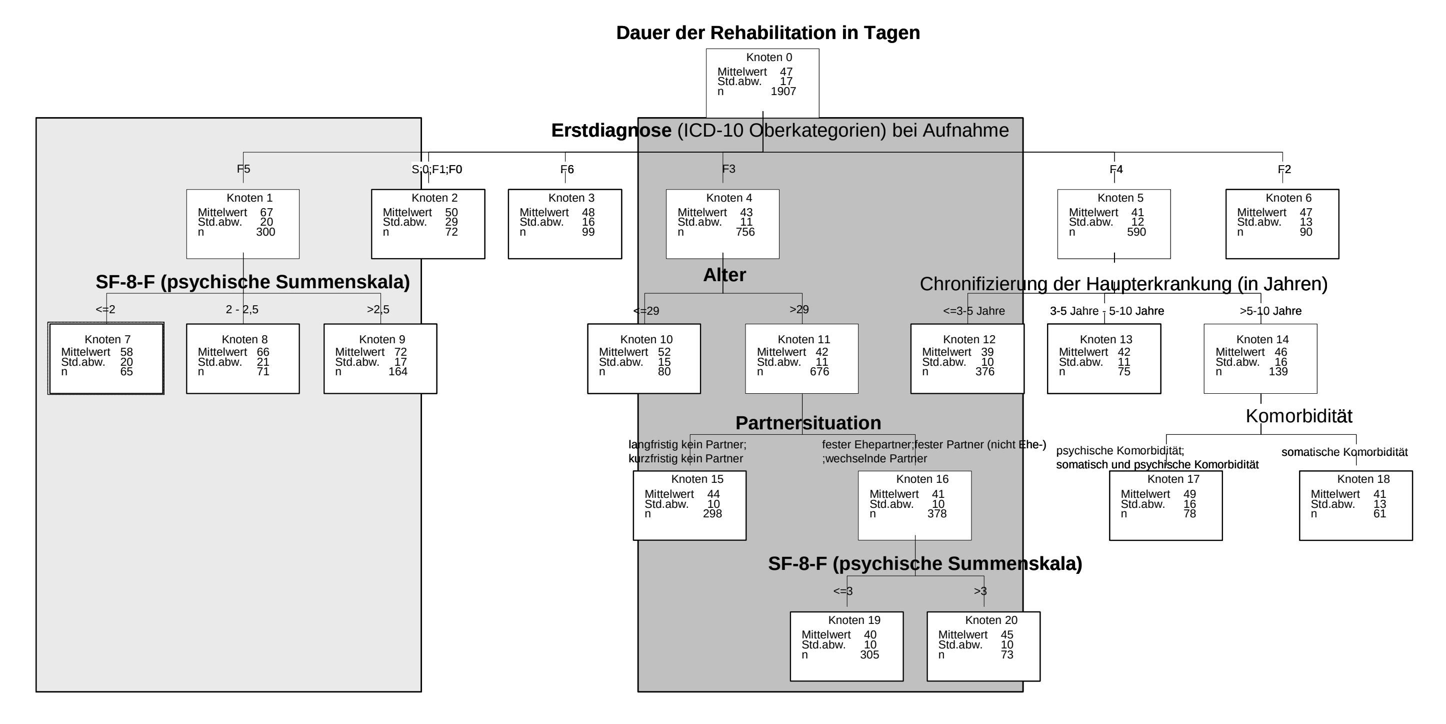



Abbildung 47: Entscheidungsbaum des „Prädiktoren-Kern-Sets“ mit dem zusätzlichen Prädiktor „Psychische Summenskala der SF-8-F“. Es wurden 14 Fallgruppen mit einer Gesamtvarianzaufklärung von 31.76% extrahiert.

Die **Ergebnisse der beiden Multiplen Regressionsanalysen** sind in den Tabelle 37 und Tabelle 38 aufgeführt. Im Gegensatz zur Position der Merkmale in der Regression-Tree-Analyse erklären die beiden Prädiktoren in der Multiplen Regressionsanalyse jeweils unmittelbar nach den Erstdiagnosen F3, F4, F6 und F2 einen bedeutsamen Varianzanteil an der Behandlungsdauer auf. Hinsichtlich der Höhe der standardisierten Regressionskoeffizienten zeigen sich für beide Prädiktoren vergleichbare Werte, wobei für die „Psychische Summenskala der SF-8-F" ein tendenziell höherer Betrag verzeichnet werden kann (stand. ß-Gewicht für die „Psychische Summenskala der SF-8-F" = .154 vs. stand. ß-Gewicht für den „Gesamtscore der HoNOS-D" = .135). Bei beiden Merkmalen handelt es sich um signifikante Prädiktoren der Vorhersage zur Behandlungsdauer.

**Testet** man nun die beiden **nicht-standardisierten Regressionskoeffizienten** anhand der Berechnungsvorschriften aus Abschnitt 9.6.2.2 gegeneinander, dann erhält man für die kritische Differenz einen Wert von = 0.951 ($Standardfehler_{SF\text{-}8\text{-}F}$ = 0.345, $Standardfehler_{HoNOS\text{-}D}$ = 0.341). Die empirische Differenz zwischen den Regressionskoeffizienten der „Psychischen Summenskala der SF-8-F" und des „Gesamtscores der HoNOS-D" beträgt 0.226. Damit ist $Diff_{emp} < Diff_{krit}$, d.h. die beiden Regressionskoeffizienten der Prädiktoren unterscheiden sich statistisch nicht signifikant voneinander.

Hinsichtlich der **erklärten Varianzaufklärung ($R^2_{adj}$) der Multiplen Regressionsanalysen** lässt sich zeigen, dass erwartungsgemäß beide Analysen zu vergleichbaren Ergebnissen gelangen, wobei die Analyse mit dem Gesamtscore der HoNOS-D eine leicht höhere Ausprägung in der Varianzaufklärung erzielt, wie auch bereits im Regression-Tree-Verfahren gezeigt werden konnte ($R^2_{adj\ HoNOS\text{-}D}$ = 33.6% vs. $R^2_{adj\ SF\text{-}8\text{-}F}$ = 33.3%).

Tabelle 37: Standardisierte und nicht-standardisierte Regressionskoeffizienten, t-Werte, Signifikanzniveaus und Toleranzwerte des „Prädiktoren-Kern-Sets“ mit dem zusätzlichen Prädiktor „Gesamtscore der HoNOS-D“ zur Vorhersage der Behandlungsdauer in absteigender Reihenfolge der standardisierten ß-Gewichte (N = 1.907)

| Prädiktoren | B | Standardisierte ß-Gewichte | t-Wert | p-Wert | Toleranz |
|---|---|---|---|---|---|
| Intercept | | | | .000 | |
| F3* | -16,72 | -0,470 | -3,66 | 0,0001 | 0,02 |
| F4* | -17,64 | -0,468 | -3,84 | 0,000 | 0,02 |
| F6* | -15,39 | -0,196 | -3,23 | 0,001 | 0,09 |
| F2* | -14,83 | -0,181 | -3,12 | 0,002 | 0,10 |
| **HoNOS-D (Gesamt)** | **2.343**[+] | **0,135** | **6,87** | **.000** | **0,91** |
| Partnersituation* (langfristig kein Partner) | 4,28 | 0,102 | 4,72 | .000 | 0,74 |
| S* | -11,30 | -0,101 | -2,25 | 0,02 | 0,171 |
| Alter | -0,10 | -0,081 | -3,19 | 0,002 | 0,54 |
| F1* | -17,00 | -0,080 | -2,82 | 0,005 | 0,43 |
| Ambulante Psychotherapie* (6-25 Termine) | 3,08 | 0,074 | 3,51 | 0,001 | 0,78 |
| F5* | 3,24 | 0,068 | 0,70 | 0,48 | 0,04 |
| Chronifizierung | 0,82 | 0,067 | 3,42 | 0,001 | 0,90 |
| Partnersituation* (kurzfristig keinen Partner) | 3,56 | 0,065 | 3,09 | 0,002 | 0,79 |
| Ambulante Psychotherapie* (Missing) | 3,16 | 0,062 | 3,16 | 0,002 | 0,90 |
| Partnersituation* (fester Partner) | 2,86 | 0,056 | 2,64 | 0,01 | 0,78 |
| Komorbidität* (som.) | -2,28 | -0,056 | -2,54 | 0,01 | 0,72 |
| Ambulante Psychotherapie* (mehr als 25 Termine) | 2,65 | 0,049 | 2,36 | 0,02 | 0,81 |
| Partnersituation* (Missing) | 1,49 | 0,030 | 1,53 | 0,13 | 0,91 |
| F0* | -15,63 | -0,030 | -1,41 | 0,16 | 0,82 |
| Ambulante Psychotherapie* (1-5 Termine) | 1,42 | 0,029 | 1,39 | 0,16 | 0,82 |
| Komorbidität* (som. und psych.) | -0,91 | -0,015 | -0,76 | 0,45 | 0,87 |
| Komorbidität* (Missing) | 0,38 | 0,010 | 0,49 | 0,63 | 0,89 |
| Partnersituation* (wechselnde Partner) | 0,17 | 0,001 | 0,07 | 0,95 | 0,89 |
| R | .587 | | | | |
| R² | .344 | | | | |
| R²adj | .336 | | | | |
| F | 42.97 | | | | |
| df | 23/1883 | | | | |
| p | .000 | | | | |

**Anmerkungen:** * dummycodierte Variablen; + z-standardardisierte Skalenwerte; S = Somatische Komorbidität

Tabelle 38: Standardisierte und nicht-standardisierte Regressionskoeffizienten, t-Werte, Signifikanzniveaus und Toleranzwerte des „Prädiktoren-Kern-Sets“ mit dem zusätzlichen Prädiktor „Psychische Summenskala der SF-8-F“ zur Vorhersage der Behandlungsdauer in absteigender Reihenfolge der standardisierten ß-Gewichte (N = 1.907)

| Prädiktoren | B | Standardisierte ß-Gewichte | t-Wert | p-Wert | Toleranz |
|---|---|---|---|---|---|
| Intercept | 57.15 | | 11.71 | .000 | |
| F3* | -15,85 | -0,445 | -3,47 | 0,001 | 0,02 |
| F4* | -16,13 | -0,428 | -3,52 | 0,001 | 0,02 |
| F6* | -13,42 | -0,171 | -2,82 | 0,005 | 0,09 |
| F2* | -13,37 | -0,163 | -2,82 | 0,005 | 0,10 |
| **SF-8-F** | **2,569**[+] | **0,145** | **7,46** | **.000** | **0,89** |
| F5* | 5,49 | 0,115 | 1,19 | 0,23 | 0,04 |
| Partnersituation* (langfristig kein Partner) | 4,32 | 0,103 | 4,78 | .000 | 0,74 |
| Alter | -0,10 | -0,079 | -3,10 | 0,002 | 0,54 |
| S* | -8,53 | -0,076 | -1,70 | 0,09 | 0,17 |
| Chronifizierung | 0,91 | 0,074 | 3,81 | 0,000 | 0,91 |
| Ambulante Psychotherapie* (6-25 Termine) | 2,98 | 0,072 | 3,40 | 0,002 | 0,78 |
| F1* | -14,31 | -0,068 | -2,37 | 0,02 | 0,43 |
| Partnersituation* (kurzfristig kein Partner) | 3,60 | 0,066 | 3,14 | 0,002 | 0,79 |
| Partnersituation* (fester Part-ner) | 3,03 | 0,059 | 2,81 | 0,005 | 0,78 |
| Ambulante Psychotherapie* (Missing) | 2,98 | 0,059 | 2,99 | 0,003 | 0,90 |
| Ambulante Psychotherapie (mehr als 25 Termine) | 2,60 | 0,048 | 2,32 | 0,02 | 0,81 |
| Komorbidität (som.) | -1,70 | -0,042 | -1,89 | 0,06 | 0,70 |
| Partnersituation* (Missing) | 1,62 | 0,033 | 1,67 | 0,10 | 0,91 |
| F0* | -11,77 | -0,022 | -1,07 | 0,29 | 0,82 |
| Ambulante Psychotherapie* (6-25 Termine) | 0,98 | 0,020 | 0,97 | 0,33 | 0,82 |
| Komorbidität* (som. u. psych.) | -0,79 | 0,013 | -0,66 | 0,51 | 0,87 |
| Komorbidität* (Missing) | 0,46 | 0,012 | 0,60 | 0,55 | 0,89 |
| Partnersituation* (wechselnde Partner) | 0,54 | 0,004 | 0,21 | 0,84 | 0,89 |
| R | .589 | | | | |
| $R^2$ | .347 | | | | |
| $R^2adj$ | .339 | | | | |
| F | 43.51 | | | | |
| df | 23/1883 | | | | |
| p | .001 | | | | |

**Anmerkungen:** * dummycodierte Variablen; [+] z-standardardisierte Skalenwerte; S = Somatische Komorbidität

**Zusammenfassung der Ergebnisse zur Hypothese 2c „Position im Prädiktoren-Kern-Set“**

**Hypothese 2c:** Es wird erwartet, dass der „Gesamtscore der HoNOS-D“ innerhalb eines Prädiktoren-Kern-Sets“ mittels Regression-Tree-Verfahren extrahiert wird und einen höheren Beitrag zur Varianzaufklärung liefert als ein vergleichbarer Fremdeinschätzungsscore („Psychische Summenskala der SF-8-F“).

Die Hypothese 2c lässt sich nur zum Teil bestätigen. Die Ergebnisse des Regression-Tree-Verfahrens konnten zeigen, dass der „Gesamtscore der HoNOS-D“ innerhalb eines „Prädiktoren-Kern-Sets“ ein bedeutsamer Prädiktor der Behandlungsdauer ist. Es konnte nicht gezeigt werden, dass der „Gesamtscore der HoNOS-D“ einen höheren Beitrag zur Varianzaufklärung liefert als die „Psychische Summenskala der SF-8-F“, weswegen die Hypothese 2c für diesen Teilaspekt verworfen werden muss.

## 10.4 Zur Bedeutung von Selbst- oder Fremdeinschätzungsinstrumenten innerhalb der Fallgruppenbildung (explorative Nebenfragestellung)

Im Folgenden soll abschließend noch eine explorative Nebenfragestellung bearbeitet werden, die den Vergleich zwischen Selbst- und Fremdeinschätzungsinstrumenten bei der Fallgruppenbildung verfolgt. Dazu wurden analog zum bisherigen Vorgehen zunächst ein Regression-Tree-Verfahren und anschließend zur Bestimmung der Bedeutsamkeit der Prädiktoren eine Multiple Regressionsanalyse gerechnet.

Zur Beantwortung der interessierenden Fragestellung wurde nun statt den fremd eingeschätzten Prädiktoren („Gesamtscore der HoNOS-D“ und „Psychische Summenskala der SF-8-F“), der selbst eingeschätzte Prädiktor „Psychische Summenskala der SF-8-S“ zum „Prädiktoren-Kern-Set“ hinzugefügt. Abbildung 48 zeigt die Ergebnisse des Regression-Tree-Verfahrens.

Wie aus Abbildung 48 zu entnehmen ist, wird der Prädiktor **„Psychische Summenskala der SF-8-S“** in *zwei Diagnosegruppen* von Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung extrahiert. So nimmt er eine bedeutende Position bei Patienten mit „Affektiven Störungen“ und bei Patienten mit „Neurotischen-, somatoformen und Belastungsstörungen“ ein. Vergleichbar mit dem Prädik-

tor „Gesamtscore HoNOS-D“ wird die Variable „Psychische Summenskala der SF-8-S“ auf der dritten Baumebene nach den Merkmalen „Affektive Störung“ und „Alter“ der Patienten extrahiert.

Hinsichtlich der **erzielten Varianzaufklärung** dieser **Regression-Tree-Analyse** lässt sich für den Prädiktor „Psychische Summenskala SF-8-S“ ein tendenziell niedrigerer Wert feststellen: $\text{Varianzaufklärung}_{\text{SF-8-S}}$ = 31.53% vs. $\text{Varianzaufklärung}_{\text{HoNOS-D}}$ = 31.82% und $\text{Varianzaufklärung}_{\text{SF-8-F}}$ = 31.76%).

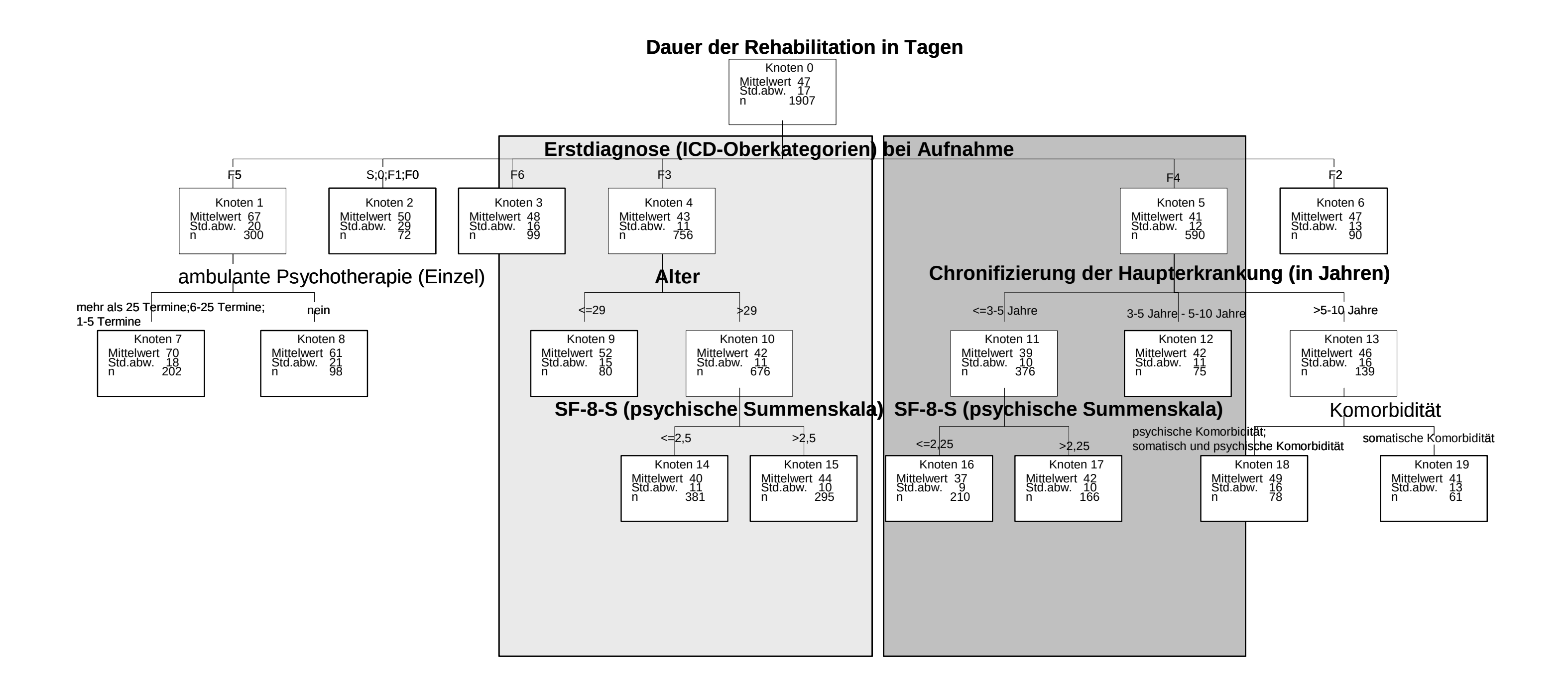



Abbildung 48: Entscheidungsbaum des „Prädiktoren-Kern-Sets“ mit dem zusätzlichen Prädiktor „Psychische Summenskala der SF-8-S“. Es wurden 13 Fallgruppen mit einer Gesamtvarianzaufklärung von 31.53% extrahiert.

Analog zum bisherigen Vorgehen wurde nun zur Überprüfung der **Bedeutsamkeit des Prädiktors „Psychische Summenskala der SF-8-S“** innerhalb des definierten „Prädiktoren-Kern-Sets“ eine **Multiple Regressionsanalyse** gerechnet. Die Ergebnisse sind der Tabelle 39 zu entnehmen.

Im Vergleich zu den beiden Analysen mit den fremd eingeschätzten Prädiktoren weist der Prädiktor „Psychische Summenskala der SF-8-S“ das niedrigste standardisierte ß-Gewicht auf ($ß\text{-Gewicht}_{SF\text{-}8\text{-}S} = 0.126$ vs. $ß\text{-Gewicht}_{SF\text{-}8\text{-}F} = 0.145$ und $ß\text{-Gewicht}_{HoNOS\text{-}D} = 0.135$) (siehe Tabelle 39).

Testet man nun wieder die nicht-standardisierten Regressionskoeffizienten gegeneinander, so erhält man für den Vergleich zwischen „Psychischer Summenskala SF-8-F“ und „Psychischer Summenskala SF-8-S“ eine $Diff_{krit} = 0.9507$ ($Standardmessfehler_{SF\text{-}8\text{-}F} = 0.345$; $Standardmessfehler_{SF\text{-}8\text{-}S} = 0.341$) und eine $Diff_{emp} = 0.38$. Damit unterscheiden sich die beiden unstandardisierten Regressionskoeffizienten also nicht signifikant voneinander. Auch der Vergleich zwischen den Prädiktoren „Psychische Summenskala der SF-8-S“ und „Gesamtscore der HoNOS-D“ erbringt hinsichtlich der Unterschiedlichkeit der unstandardisierten Regressionskoeffizienten kein signifikantes Ergebnis ($Diff_{krit} = 0.945$; $Diff_{emp} = 0.154$; $Standardmessfehler_{SF\text{-}8\text{-}S} = 0.345$, $Standardmessfehler_{HoNOS\text{-}D} = 0.341$).

Tabelle 39: Standardisierte und nicht-standardisierte Regressionskoeffizienten, t-Werte, Signifikanzniveaus und Toleranzwerte des „Prädiktoren-Kern-Sets“ mit dem zusätzlichen Prädiktor „Psychische Summenskala der SF-8-S“ zur Vorhersage der Behandlungsdauer in absteigender Reihenfolge der standardisierten ß-Gewichte (N = 1.907)

| **Prädiktoren** | **B** | **Standardisierte ß-Gewichte** | **t-Wert** | **p-Wert** | **Toleranz** |
|---|---|---|---|---|---|
| Konstante | 58.16 | | 11.89 | .000 | |
| F3* | -15,29 | -0,430 | -3,34 | 0,001 | 0,02 |
| F4* | -16,16 | -0,429 | -3,51 | 0,001 | 0,023 |
| F6* | -13,07 | -0,167 | -2,73 | 0,006 | 0,09 |
| F2* | -12,07 | -0,147 | -2,53 | 0,01 | 0,10 |
| **SF-8-S** | **2,189+** | **0,126** | **6,41** | **0,000** | **0,91** |
| F5* | 5,68 | 0,119 | 1,23 | 0,22 | 0,04 |
| Partnersituation* (langfristig kein Partner) | 4,43 | 0,109 | 4,88 | 0,000 | 0,74 |
| Alter | -0,12 | -0,092 | -3,60 | 0,000 | 0,54 |
| S* | -8,39 | -0,075 | -1,66 | 0,10 | 0,17 |
| Chronifizierung | 0,90 | 0,073 | 3,72 | 0,000 | 0,91 |
| Partnersituation* (kurzfristig kein Partner) | 3,52 | 0,064 | 3,06 | 0,002 | 0,79 |
| F1* | -13,57 | -0,064 | -2,24 | 0,03 | 0,43 |
| Ambulante Psychotherapie* (6-25 Termine) | 2,60 | 0,063 | 2,93 | 0,003 | 0,77 |
| Ambulante Psychotherapie* (Missing) | 3,16 | 0,062 | 3,16 | 0,002 | 0,90 |
| Komorbidität* (som.) | -2,40 | -0,059 | -2,68 | 0,007 | 0,72 |
| Partnersituation* (fester Partner) | 3,00 | 0,059 | 2,77 | 0,006 | 0,78 |
| Ambulante Psychotherapie* (mehr als 25 Termine) | 2,10 | 0,0387 | 1,86 | 0,06 | 0,81 |
| Partnersituation* (Missing) | 1,42 | 0,0285 | 1,46 | 0,15 | 0,91 |
| Komorbidität* (som. u. psych.) | -1,18 | -0,020 | -0,99 | 0,32 | 0,87 |
| F0 | -9,56 | -0,018 | -0,86 | 0,39 | 0,83 |
| Ambulante Psychotherapie* (1-5 Termine) | 0,72 | 0,015 | 0,70 | 0,493 | 0,81 |
| Komorbidität* (Missing) | 0,27 | 0,007 | 0,35679798 | 0,72 | 0,89 |
| Partnersituation* (wechselnde Partner) | 0,37 | 0,003 | 0,14311648 | 0,89 | 0,89 |
| R | .585 | | | | |
| R² | .342 | | | | |
| R²adj | .334 | | | | |
| F | 42.57 | | | | |
| df | 23/1883 | | | | |
| p | .001 | | | | |

Anmerkung: * dummycodierte Variablen; + z-standardardisierte Skalenwerte; S = Somatische Komorbidität

## 11 Diskussion

Die **differenzierte Erfassung des Schweregrades** nimmt in vielen verschiedenen Fragestellungen zur Erforschung der Behandlung von Patienten mit psychischen Störungen eine zentrale Rolle ein. So gewinnt sie auch im Zusammenhang mit ökonomischen Gesichtspunkten, insbesondere der Inanspruchnahmefrequenz von Krankenhausaufenthalten und der Behandlungsdauer, zunehmend an Einfluss (Borgart & Meermann, 1999; Roick et al., 2002a; Roick et al., 2002b). Ihre Erhebung basiert meist auf Selbsteinschätzungen durch die Patienten, aber auch auf Fremdeinschätzungen durch andere Personen (vorwiegend Therapeuten; vgl. hierzu ausführlich Kapitel 5). Dabei kommen Selbst- und Fremdeinschätzungen komplementäre Funktionen zu: Z.B. sind Fremdeinschätzungen in all jenen Bereichen von Vorteil, in denen der Patient selbst keine Auskunft geben kann. Auch innerhalb von Klassifikationssystemen bei Patienten mit psychischen Störungen in stationärer psychotherapeutischer Behandlung stellt die differenzierte Erfassung des Schweregrades eine der zentralen Dimensionen zur Vorhersage des Ressourcenverbrauchs dar. Vergleicht man international für den Bereich psychiatrischer/psychotherapeutischer Versorgung entwickelte Patientenklassifikationssysteme miteinander, die in ihrer Konzeption über allgemeine pauschalierte Entgeltsysteme hinausgehen, so zeigt sich u.a. die Relevanz eines Instrumentes: Der „Health of the Nation Outcome Scales (HoNOS)“ (Buckingham et al., 1998; Carthew et al., 2003; Gaines et al., 2003). Obwohl bereits einige Fremdeinschätzungsinstrumente im deutsprachigen Raum existieren, besteht derzeit noch ein erhebliches Defizit hinsichtlich Instrumente zur differenzierten Erfassung des Schweregrades (vgl. hierzu Kapitel 5, Abschnitt 5.2). Deswegen wurde der HoNOS im Rahmen eines mehrstufigen, konsensus-gestützten Prozesses von der Arbeitsgruppe für Psychotherapie- und Versorgungsforschung (AGPV) am Institut für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf in die deutsche Sprache übersetzt (vgl. hierzu Andreas et al., 2004b). Zahlreiche empirische Untersuchungen zur Originalversion zeigen allerdings auf, dass die Ergebnisse hinsichtlich psychometrischer Eigenschaften der HoNOS heterogen sind.

In der vorliegenden Arbeit wurde die Eignung der deutschen Version der HoNOS (HoNOS-D) zur differenzierten Erfassung der Schwere einer psychischen Erkrankung innerhalb der Entwicklung eines Klassifikationssystems von Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung überprüft. In einem **ersten Schritt** wurde eine empirische Prüfung der *generellen Eignung der HoNOS-D*

hinsichtlich psychometrischer Eigenschaften, insbesondere der Praktikabilität und Reliabilität, anhand einer weitestgehend repräsentativen Stichprobe von Patienten aus 11 verschiedenen Rehabilitationsfachkliniken des Indikationsbereiches „Psychische Störungen“ vorgenommen. In einem **zweiten Schritt** erfolgten empirische Analysen zur Überprüfung der *spezifischen Eignung der HoNOS-D* innerhalb eines entwickelten Klassifikationssystems für Patienten mit psychischen Störungen. Die patienten- und therapeutenseitige Datenerhebung wurde zu Beginn der stationären psychosomatischen/psychotherapeutischen Rehabilitationsmaßnahme durchgeführt. Am Ende der Behandlung wurden die Angaben zur Behandlungsdauer der Patienten erfasst. Das Studiendesign lässt sich damit als prospektive Feldstudie mit Zwei-Punkt-Erhebung charakterisieren.

Im Folgenden sollen nun die Ergebnisse der empirischen Analysen getrennt für die beiden Hauptfragestellungen diskutiert werden. Dabei soll im Abschnitt 11.1 auf die Ergebnisse zur empirischen Überprüfung der **generellen Eignung der HoNOS-D**, insbesondere der Praktikabilität und Reliabilität eingegangen werden. Darauf folgend sollen im Abschnitt 11.2 die Ergebnisse zur empirischen Prüfung der **spezifischen Eignung der HoNOS-D** innerhalb des entwickelten Klassifikationssystems für Patienten mit psychischen Störungen zusammenfassend diskutiert werden. In Abschnitt 11.3 soll die **Generalisierbarkeit** der empirischen Befunde vorliegender Arbeit erörtert werden, bevor abschließend Vorschläge zur **Weiterentwicklung** und **Empfehlungen zum Einsatz** der HoNOS-D innerhalb eines nationalen Patientenklassifikationssystems gegeben werden (Abschnitt 11.4 und 11.5). Der damit einhergehende **weitere Forschungsbedarf** wird in Abschnitt 11.6 skizziert.

## 11.1 Empirische Überprüfung der generellen Eignung der HoNOS-D

Mindestanforderungen von Instrumenten innerhalb eines Klassifikationssystems sollten sich an deren psychometrische Eigenschaften, insbesondere die Praktikabilität und Reliabilität, richten. Im Folgenden sollen diese beiden Aspekte der empirischen Überprüfung der HoNOS-D vor dem Hintergrund der Hypothesen 1a bis 1e diskutiert werden.

### *11.1.1 Praktikabilität*

Von besonderer Bedeutung bei der Konstruktion eines Fremdeinschätzungsinstrumentes ist die Praktikabilität und somit die Eignung zum routinemäßigen Einsatz zur Datenerfassung.

Bei der Entwicklung der HoNOS (Wing et al., 1998) wurde deswegen besonderer Wert auf diesen Aspekt gelegt.

In Hypothese 1a war postuliert worden, dass 9 der 12 Items und der Gesamtscore der deutschen Version der HoNOS einen niedrigen Anteil fehlender Werte (unter 10%) aufweisen sollten. Dagegen wurde für die Items 8 („Andere Symptome"), 11 („Probleme durch die Wohnbedingungen") und 12 („Probleme im Beruf und in der Freizeit") der HoNOS-D ein höherer prozentualer Anteil fehlender Werte (über 10%) erwartet.

Erwartungsgemäß zeigte sich für die 9 der 12 Items der HoNOS-D ein niedriger prozentualer Anteil fehlender Werte (Spannweite von 1.3% bis 3.9%). Auch der Gesamtscore der HoNOS-D wies mit 3.4% fehlender Werte ein zufrieden stellendes Ergebnis auf. Entgegen der Hypothese 1a, wonach für 3 der 12 Items ein höherer prozentualer Anteil fehlender Werte postuliert worden war, zeigten sich auch hier befriedigende Ergebnisse mit Fehlquoten unter 10%.

Bei Betrachtung der Einzelitemebene fällt auf, dass **Item 8 „Andere Symptome"** im Vergleich zu den anderen Items der HoNOS-D mit 8.6% den höchsten prozentualen Anteil fehlender Werte aufweist. Dieses Resultat geht auch mit empirischen Befunden anderer Studien konform (Buckingham et al., 1998; Trauer, 2003). Als Erklärung für dieses Ergebnis können verschiedene Gründe in Frage kommen. Zum einen wurde das Item 8 von vornherein sehr heterogen konzipiert, um zehn verschiedene *andere psychische oder verhaltensbezogene Probleme* erfassen zu können. Des Weiteren wird im Rahmen der Beurteilerinstruktion darauf hingewiesen, dasjenige Problem auszuwählen, welches bei dem Patienten im größten Ausmaß vorhanden ist. D.h. dass der Beurteiler gebeten wird, eines dieser zehn genannten Probleme zu fokussieren. Hinzu kommt, dass im Glossar, im Gegensatz zu den anderen 11 Items der HoNOS-D, keine differenzierten Ankerbeispiele für die zehn weiteren Probleme aufgeführt sind.

Mit dem vorliegenden Studiendesign könnten nun mehrere Schwierigkeiten verbunden gewesen sein. So wäre es denkbar, dass aufgrund der Heterogenität und damit der Vielzahl der Antwortmöglichkeiten die Auswahl eines Problems schwierig war, zumal keine differenzierten Operationalisierungen im Glossar vorhanden waren. Es wäre aber auch denkbar, dass Unklarheiten bezüglich der Auswahl *eines* der zehn genannten Probleme bestanden und die Beurteiler mehrere Angaben machten, was letztlich als fehlender Wert kodiert wurde. An dieser Stelle ist zu betonen, dass im Rahmen des realisierten Studiendesigns keine Schulung der Ra-

ter im HoNOS-D durchgeführt wurde. Da aber mit dem Item 8, besonders für Patienten mit psychischen Störungen in stationärer psychotherapeutischer Behandlung, eine hohe klinische Relevanz verbunden ist (vgl. hierzu auch Audin et al., 2001), soll an dieser Stelle auf einen konzeptionellen Überarbeitungsbedarf des Items 8 der HoNOS-D hingewiesen werden. Dieser Aspekt wird in Abschnitt 11.4 näher fokussiert.

Entgegen der Hypothese 1a zeigte sich weiterhin, dass die Items 11 „Probleme durch die Wohnbedingungen" und 12 „Probleme im Beruf und in der Freizeit" einen niedrigeren prozentualen Anteil fehlender Werte als erwartet aufwiesen. Auch im MH-CASC-Projekt konnten in diesen Items über das gesamte Setting (stationäre Patienten und ambulante Patienten) weniger als 10% fehlender Werte verzeichnet werden (Buckingham et al., 1998), wenngleich im stationären Setting im MH-CASC-Projekt und in der Studie von Trauer et al. (1999) für diese beiden Items mehr als 10% fehlender Werte festgestellt wurden (Buckingham et al., 1998). Obwohl für eine adäquate Beurteilung dieser Items Informationen aus dem häuslichen und sozialen Umfeld der Patienten nötig waren und der Beurteilungszeitraum mit einer Zeitspanne von 7 Tagen eher kurz gewählt war, schienen die Rater für eine Einschätzung ausreichend Informationen besessen zu haben. Damit kommen vermutlich gleichzeitig settingspezifische Anteile zum Ausdruck. So besteht eine Zielsetzung der stationären Rehabilitationsbehandlung von Patienten mit psychischen Störungen aus Sicht der Rentenversicherungsträger im Erhalt bzw. der Wiederherstellung der Erwerbstätigkeit. Dagegen sieht die Gesetzliche Krankenversicherung eine gesundheitliche und soziale Integration der Patienten im Vordergrund der Therapie (Schulz & Koch, 2002). Beide Aspekte finden sich nun operationalisiert in den Items 11 und 12 der HoNOS-D wieder, weswegen auch geschlussfolgert werden kann, dass diese beiden Items therapierelevante Informationen beinhalten und daher bereits zu Beginn der Behandlung eruiert werden.

Ergänzend zu den deskriptiven Itemanalysen sollen nun noch erste Ergebnisse einer **explorativen Befragung** von Therapeuten aus der Rehabilitationsbehandlung Suchtkranker hinsichtlich der **Praktikabilität** der HoNOS-D diskutiert werden. Zunächst gaben die Befragten und im HoNOS-D bereits geübten Behandler (n = 7) eine durchschnittliche Bearbeitungsdauer von 5 Minuten für die Einschätzungen an, die auch im Einklang mit anderen Studien steht (Wing et al., 1998; Orrell et al., 1999; Browne et al., 2000). Damit stellt die Bearbeitungszeit eine in der Routineanwendung vertretbare Größe dar. Weiterhin ist hervorzuheben, dass die Mehrzahl der Beurteiler alle nötigen Informationen aus dem Aufnahmegespräch mit dem Pa-

tienten erhielten. Auch dieser Befund gibt erste Hinweise auf die Praktikabilität der HoNOS-D. Die empirischen Befunde der deskriptiven Analyse fehlender Werte untermauernd, zeigte sich auch in der direkten Therapeutenbefragung, dass die Items 8 („Andere Symptome“), 10 („Probleme mit alltäglichen Aktivitäten“), 11 („Probleme durch die Wohnbedingungen“) und 12 („Probleme im Beruf und in der Freizeit“) noch mit den größten Beurteilungsschwierigkeiten behaftet waren. Bei der Darstellung dieser ersten Befunde ist einschränkend anzumerken, dass es sich zunächst nur um eine sehr kleine Stichprobe von Beurteilern handelt, weswegen kein Anspruch auf Generalisierbarkeit bzw. Repräsentativität der Ergebnisse besteht. Die gerade erwähnten Aspekte (u.a. Beurteilungsdauer in der Routine bzw. Erhalt der Informationen für ein Rating) sollen an größeren Stichproben im Rahmen von Schulungsangeboten zum HoNOS-D weiter untersucht werden.

Vorrangiges Ziel der Testautoren der HoNOS (Wing et al., 1996; Wing et al., 1998) war die Konstruktion eines Instrumentes, welches in der Routine verwendet werden sollte. Dieser Aspekt konnte in der vorliegenden Untersuchung bestätigt werden, wobei insgesamt die Aussage getroffen werden kann, dass es sich bei der HoNOS-D um ein praktikables Instrument handelt. Dabei soll aber auch noch darauf hingewiesen werden, dass Item 8, welches für den vorliegenden Untersuchungsbereich hohe klinische Relevanz besitzt, erheblichen Optimierungsbedarf aufweist.

### *11.1.2 Verteilungseigenschaften*

In der Hypothese 1b war die Erwartung formuliert worden, dass 2 der 12 Items (Item 3 „Substanzmissbrauch“ und Item 6 „Halluzinationen und Wahnvorstellungen“) aufgrund der erwarteten Stichprobencharakteristika rechts schiefe Verteilungseigenschaften aufweisen sollten. Für alle anderen Items und den Gesamtscore waren moderate bis zufrieden stellende Zellbesetzungen in den einzelnen Kategorien erwartet worden.

Die Hypothese 1b kann aufgrund der empirischen Befunde der deskriptiven Analyse nur zum Teil bestätigt werden. Hypothesenkonform wiesen **Item 3 „Substanzmissbrauch“** und **Item 6 „Halluzinationen und Wahnvorstellungen“** deutlich rechts schiefe Verteilungen auf. Entgegen der Erwartung, dass Item 2 „Selbstverletzung“ moderate bis zufrieden stellende Zellbesetzungen in den möglichen Kategorien aufweisen sollte, fand sich auch hier eine rechts schiefe Verteilung. Diese Ergebnisse stehen damit im Einklang mit den empirischen Befun-

den anderer Studien (Orrell et al., 1999; Trauer et al., 1999; Brooks, 2000), besonders ist hier die Studie von Audin et al. (2001) hervorzuheben. Die Autoren überprüften anhand einer konsekutiv erhobenen Stichprobe von Patienten in ambulanter psychotherapeutischer Behandlung in Großbritannien die Verteilungseigenschaften der Items der finalen HoNOS Version. Auch sie fanden nur moderate Verteilungseigenschaften für die Mehrzahl der Items, darunter auch die Items 2, 3 und 6 (Audin et al., 2001). Das legt die Vermutung nahe, dass die Verteilungseigenschaften der Items im HoNOS-D von der zugrunde liegenden Stichprobe bzw. dem Patientenklientel abhängen. Die Stichprobe der vorliegenden Studie ist dadurch gekennzeichnet, dass Patienten mit einer F1-Diagnose nach ICD-10 („Psychische und Verhaltensstörungen durch psychotrope Substanzen“) nur zu einem Prozentsatz von 7.1% und Patienten mit einer F2-Diagnose („Schizophrenie, schizotype und wahnhafte Störungen“) (ICD-10, Dilling et al., 1993) nur mit einem prozentualen Anteil von 5.4% vorkamen. Daraus kann wiederum abgeleitet werden, dass diejenigen Patienten, die vermutlich in den Items 3 und 6 Probleme aufweisen sollten, nur zu einem geringen Anteil in der vorliegenden Untersuchungsstichprobe vertreten waren. Damit ergäbe sich weiterer Forschungsbedarf die Verteilungseigenschaften von Item 3 und Item 6 anhand einer Stichprobe mit jeweils höherem Anteil an den aufgeführten Diagnosen zu prüfen.

Die Befunde der deskriptiven Itemanalyse zeigen ebenfalls deutlich rechts schiefe Verteilungseigenschaften für das **Item 2 „Selbstverletzung“**. Auch dieses Ergebnis legt die Vermutung nahe, dass Patienten in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung eher weniger unter selbstverletzenden oder suizidalen Verhalten leiden. Ein weiterer Aspekt betrifft unter Umständen aber auch die ungenaue Betitelung des Items: So ist das Item 2 im Ratingbogen lediglich mit der Überschrift „Absichtliche Selbstverletzung“ aufgeführt. Item 2 umfasst aber sowohl die Suizidgefährdung als auch das selbstverletzende Verhalten von Patienten. Ein weiteres Erklärungsmodell für die moderaten Verteilungseigenschaften könnte somit auch die **fehlende Schulung der Rater** einbeziehen. Aufgrund der empirischen Ergebnisse von Rock und Preston (2001), die in einem Experimentaldesign aufzeigen konnten, dass auch durch Schulungen keine höheren Werte zur Interraterreliabilität erzielt werden konnten, wurde in der vorliegenden Studie auf eine Schulung der Rater verzichtet. Darin könnte aber vermutlich auch ein Grund für die moderaten Verteilungseigenschaften von 3 der 12 Items liegen.

Hypothesenkonform konnten für die übrigen **9 der 12 Items** und für den **Gesamtscore der HoNOS-D** moderate bis zufrieden stellende Verteilungseigenschaften nachgewiesen werden. Dabei wiesen 6 der 12 Items zufrieden stellende Zellbesetzungen in den fünf Kategorien der HoNOS-D auf. Bezüglich der Verteilungseigenschaften für den Gesamtscore der HoNOS-D konnten befriedigende Schiefe- und Kurtosisindizes gezeigt werden.

Bei differenzierter Betrachtung der Verteilungseigenschaften der Items der HoNOS-D lässt sich nun **ableiten**, dass sich für die vorliegende Untersuchungsstichprobe von Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung 3 der 12 Items als nicht relevant zur Erfassung der Schwere der Erkrankung erwiesen haben (Item 2, 3 und 6). Für 3 weitere Items (Item 1 „Aggression“, 5 „Probleme mit körperlichen Erkrankungen“ und 11 „Probleme durch die Wohnbedingungen“) zeigte sich aufgrund ihres eingeschränkten Streubereiches nur eine bedingte Relevanz. Nur 6 der 12 Items erwiesen sich als unmittelbar relevant für die differenzierte Erfassung der Schwere einer psychischen Störung von Patienten in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung. In diesem Zusammenhang muss aber auch festgehalten werden, dass ursprünglich der HoNOS vornehmlich für schwerstkranke Patienten in stationärer psychiatrischer/psychotherapeutischer Behandlung konzipiert wurde, aber auch mit dem Anspruch in allen Versorgungsbereichen psychisch Kranker einsetzbar zu sein (Wing et al., 1998). Deswegen sollten die Verteilungseigenschaften und die postulierte Stichprobenabhängigkeit in weiteren Studien an verschiedenen Patientengruppen aus den unterschiedlichsten Versorgungsbereichen psychisch Kranker untersucht werden. **Abschließend** lässt sich nochmals festhalten, dass sich der Gesamtscore der HoNOS-D hinsichtlich seiner Verteilungseigenschaften gegenüber den Einzelitems als überlegen erwies.

### *11.1.3 Item-Interkorrelationen*

Hypothesenkonform (Hypothese 1c) zeigte sich hinsichtlich der **Interkorrelationsmatrix** der Items der HoNOS-D nur *eine* signifikante Korrelation mindestens mittlerer Effektstärke. Alle anderen Korrelationen blieben unter einem Wert von .30. Die von den Testautoren (Wing et al., 1996) berichteten Interkorrelationen zwischen Item 4 „Kognitive Probleme“ und Item 10 „Probleme mit alltäglichen Aktivitäten“ sowie zwischen Item 2 „Selbstverletzung“ und Item 7 „Gedrückte Stimmung“ konnten in der vorliegenden Untersuchung nicht aufgezeigt werden.

Es ergab sich eine Korrelation zwischen Item 7 „Gedrückte Stimmung“ mit dem Item 10 „Probleme mit alltäglichen Aktivitäten“, womit weiterer Forschungsbedarf verbunden ist.

Die Testautoren konzipierten das Instrument mit der Zielsetzung, dass jedes Item für sich genommen ein distinktes Merkmal sein sollte und betrachteten die 12 Items als weitest gehend unabhängig voneinander (Wing et al., 1996). Die postulierte weitestgehende Unabhängigkeit der 12 Items konnte hier repliziert werden. Ausgehend von dieser Konzeption wurde in der vorliegenden Untersuchung auf eine explizite Überprüfung der ergänzend vorgeschlagenen Skalenstruktur verzichtet. Weitergehende Analysen sollten dennoch prüfen, in wie weit sich die Skalen und die Skalenstruktur replizieren lassen. Für die vorliegende Untersuchung lässt sich festhalten, dass jedes einzelne Item weitestgehend unabhängig von den anderen Items ist und somit maximalen Informationsgehalt besitzt. Dieses Ergebnis spiegelt sich auch in der berichteten moderaten internen Konsistenz des Gesamtscores der HoNOS-D wider (Cronbach´s alpha = .64).

### *11.1.4 Interraterreliabilitäten*

Anhand zweier ergänzend konsekutiv erhobener Stichproben aus der Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie (insgesamt N = 37, Teilstichprobe 1) und der Rehabilitationsbehandlung Suchtkranker (N = 34, Teilstichprobe 2) wurden **Interraterreliabilitäten** zwischen jeweils zwei Beurteilern für die Items der HoNOS-D bestimmt. Auch bei diesen Ergebnissen soll vorangestellt werden, dass die Beurteiler noch keine Schulung im HoNOS-D erhalten hatten. Die Ergebnisse sollen nun zusammenfassend für die beiden Teilstichproben vor dem Hintergrund der Hypothese 1d diskutiert werden.

In der Hypothese 1d wurden für die Items 4 „Kognitive Probleme“, 8 „Andere Symptome“, 10 „Probleme mit alltäglichen Aktivitäten“, 11 „Probleme durch die Wohnbedingungen“ und 12 „Probleme im Beruf und in der Freizeit“ niedrige Übereinstimmungen in den Koeffizienten bzw. niedrige prozentuale Übereinstimmungen erwartet. Für die übrigen Items und den Gesamtscore sollten sich dagegen befriedigende Werte in den Koeffizienten zur Beurteilung der Interraterreliabilität bzw. befriedigende prozentuale Übereinstimmungen zeigen.

Die Analysen zur Interraterreliabilität ergaben für die beiden Teilstichproben hinsichtlich der einzelnen Items der HoNOS-D heterogene Befunde. Während in *Teilstichprobe 1* nur eine

befriedigende Höhe der Interraterkoeffizienten für die Items 1 „Aggression“, 3 „Substanzmissbrauch“, 4 „Kognitive Probleme“, 6 „Halluzinationen und Wahnvorstellungen“, 7 „Gedrückte Stimmung“, 9 „Probleme mit Beziehungen“ und 10 „Probleme mit alltäglichen Aktivitäten“ festgestellt werden konnten, wiesen in der *Teilstichprobe 2* alle Items eine befriedigende Interraterreliabilität auf. Damit zeigten sich in den beiden Stichproben für vier Items konsistente Befunde: Für das Item 1 „Aggression“, 3 „Substanzmissbrauch“, Item 6 „Halluzinationen und Wahnvorstellungen“ und für das Item 9 „Probleme mit Beziehungen“. Dieser Befund steht weitestgehend im Einklang mit den empirischen Befunden der Testautoren (vgl. Tabelle 8 zur Interraterreliabiltität der HoNOS-Items in Abschnitt 6.1, Kap.6, Wing et al., 1998) und mit einer weiteren Studie von Trauer et al. (1999). In der Studie der Testautoren wies das Item 3 mit ICC = .61 keine befriedigende Höhe auf, für Item 6 konnten aber in den zitierten Studien zufrieden stellende Kennwerte erzielt werden. Dabei sind die empirischen Befunde der anderen Studien nur bedingt mit den hier berichteten Ergebnissen vergleichbar, da in der vorliegenden Untersuchung die Beurteiler noch *keine Schulung* erhalten hatten. Es lässt sich jedoch aussagen, dass das Item 6 anscheinend in befriedigendem Ausmaß im Glossar operationalisiert wurde und sich auch ohne Schulung reliable Werte erzielen lassen.

In der *Teilstichprobe 1* wurden unbefriedigende Interraterreliabilitäten für die Items 2, 5, 8, 11 und 12 ermittelt. Hypothesenkonform und im Einklang mit anderen empirischen Befunden (Wing et al., 1998; Trauer et al., 1999; Brooks, 2000) kann damit eine mangelnde Interraterreliabilität für die Items 11 und 12 konstatiert werden.

Auch hier können wieder verschiedene Gründe für die unbefriedigenden Ergebnisse in Frage kommen. Als Erstes soll auf die relativ geringen Stichprobenumfänge der Teilstichprobe hingewiesen werden, angesichts derer es sich hier zunächst nur um *erste Hinweise* handeln kann und die Erhebung größerer Stichproben zur Überprüfung der Interraterreliabilität notwendig scheint (vgl. hierzu auch Kalkulation der Power in Abschnitt 10.1.2.3). Ein zweiter Aspekt betrifft sicherlich auch hier die Tatsache, dass die Beurteiler der beiden Teilstichproben noch keine Schulung erhalten hatten. Sollte der HoNOS-D allerdings in der Routineerhebung Verwendung finden, kann aus ökonomischen Gründen vermutlich keine flächendeckende Schulung der Beurteiler realisiert werden. Damit wäre die Entwicklungsaufgabe einer differenzierten Beurteilerinstruktion und eines differenzierten, auch für den Bereich von Patienten mit psychischen Störungen in stationärer Behandlung, relevanten Glossars verbunden, welches

dann in die Routineerhebung implementiert werden müsste. Dieser Aspekt soll im Abschnitt 11.4 näher ausgeführt werden. Ein weiteres statistisches und methodisches Problem könnte auch hier wieder in der mangelnden Variation der einzelnen Items der HoNOS-D zu sehen sein, insbesondere für den Bereich der stationären psychosomatischen/psychotherapeutischen Krankenhausbehandlung. Wie bereits in Abschnitt 9.6.1.1 vorliegender Arbeit ausgeführt, stellt eine ausreichende Höhe der Varianz der Beurteilereinschätzungen eine Voraussetzung zur Bestimmung des Intraklassenkoeffizienten dar.

Abschließend sollte auch noch angemerkt werden, dass eine verlässliche Bewertung und Bestimmung der Interraterreliabilitäten erst **nach einer Schulung** der Beurteiler vorgenommen werden kann. **Zusammenfassend** lässt sich damit festhalten, dass die Hypothese 1d nur zum Teil bestätigt werden konnte. In beiden Teilstichproben zeigten sich aber auch für den Gesamtscore der HoNOS-D hinsichtlich der Interraterreliabilität mit Werten von .70 und .83 (Finn-Koeffizient) zufrieden stellende Ergebnisse.

#### *11.1.5 Test-Retest-Reliabilität*

Hypothese 1e besagte, dass 8 der 12 Items sowie der „Gesamtscore der HoNOS-D“ zufrieden stellende Werte in der Test-Retest-Reliabilität aufweisen sollten. Dagegen sollten sich für die Items 1 „Aggression“, 3 „Substanzmissbrauch“, 7 „Gedrückte Stimmung“ und 10 „Probleme mit alltäglichen Aktivitäten“ eher moderate Koeffizienten zeigen (ICC < .70).

Hypothesenkonform wiesen 8 der 12 Items sowie der „Gesamtscore der HoNOS-D“ zufrieden stellende Intraklassenkoeffizienten auf, bestimmt anhand einer Subgruppe von Patienten, die sich während der stationären Behandlung hinsichtlich ihres Behandlungsergebnisses als unverändert erwiesen. Für die Items 1, 3, 7 und 10 muss die Hypothese aufgrund höherer als erwarteter verworfen werden, was überwiegend im Sinne einer guten Test-Retest-Reliabilität zu interpretieren und daher zu begrüßen ist. Auch Wing et al. (1998) fanden für 11 der 12 Items weitestgehend zufrieden stellende Intraklassenkoeffizienten zwischen .74 und .88 (außer für Item 1 „Aggression“). In der Studie von Orrell et al. (1999) wies dagegen nur das Item 4 „Kognitive Probleme“ eine befriedigende Test-Retest-Reliabilität auf, die Mehrzahl der anderen Items zeigten dagegen nur moderate Reliabilitäten. Besonders niedrige Koeffizienten fanden sich dort für die Items 1 „Aggression“, 3 „Substanzmissbrauch“, 7 „Gedrückte Stimmung“ und 11 „Probleme durch die Wohnbedingungen“, wobei die Autoren dies auf hohen

Veränderungsraten der Patienten zurückführten. Um derartige Einflüsse ausschließen zu können, wurden in der vorliegenden Untersuchung nur solche Patienten untersucht, die sich hinsichtlich des Behandlungsergebnisses nicht verändert haben.

**Zusammenfassend** kann damit festgehalten werden, dass aufgrund der empirischen Befunde zur Test-Retest-Reliabilität von einer weitestgehend reliablen Messung ausgegangen werden kann. Die Ergebnisse stehen damit aber auch im starken Kontrast zu den empirischen Befunden bezüglich der Interraterreliabilitäten, insbesondere für die Teilstichprobe 1. Kritisch ist an dieser Stelle das vorliegende Studiendesign aufzuführen, welches vorsah, den Therapeutenerhebungsbogen sowohl zu Beginn als auch am Ende der Behandlung in einem zusammenhängenden Dokument auszuhändigen. In wie weit die Therapeuten von der Möglichkeit Gebrauch gemacht haben, am Ende der Behandlung noch einmal Einsicht in den zu Beginn der Behandlung ausgefüllten Bogen zu nehmen, bleibt damit unklar. In einer weiteren Studie sollte eine Überprüfung der Test-Retest-Reliabilität erfolgen, bei der der Therapeutenbogen zu Beginn und zum Ende der Therapie separat ausgegeben wird.

### *11.1.6 Methodische Beschränkungen und Probleme*

Ein wichtiger Kritikpunkt bei der Durchführung der Untersuchung ist, dass sowohl die Beurteiler der Hauptuntersuchung als auch die Beurteiler der beiden ergänzend erhobenen Teilstichproben **keine Schulung im HoNOS-D** erhalten hatten. Obwohl eine kontrollierte Experimentalstudie aufzeigte, dass trotz Schulung im HoNOS-D keine höheren Interraterreliabilitäten erzielt werden konnten (Rock & Preston, 2001), werfen die vorliegenden empirischen Befunde die Frage nach dem Bedarf an Schulung im HoNOS-D auf. In dem vorliegenden Studiendesign des QS-Reha-Projektes erwies sich das beigefügte Glossar zur Anwendung der HoNOS-D als nicht differenziert genug und unklar bleibt auch, inwieweit die Beurteiler dieses zur Einschätzung im HoNOS-D zugrunde gelegt haben. Auch der Umstand, dass ein klinisch erfahrener Beurteiler bei der Einschätzung des Schweregrades vermutlich „seinen schwerstgestörten Patienten“ als Bezugsgröße benutzt, könnte bei der mangelnden Interraterreliabilität der Teilstichproben eine Rolle gespielt haben. Auch hier hätte es einer Schulung bedurft, die einen einheitlichen „Bezugs- bzw. Referenzrahmen“ vermittelt.

Ein zweites Problem bezieht sich auf den **Beurteilungszeitraum**. Analog zu den anderen im Rahmen der Pilotphase des Qualitätssicherungsprogramms der Gesetzlichen Krankenversi-

cherung eingesetzten Instrumenten wurde in der vorliegenden Untersuchung der Beurteilungszeitraum für die letzten 7 Tage vor Beginn der stationären psychosomatischen/psychotherapeutischen Rehabilitationsmaßnahme festgelegt. In anderen empirischen Studien lag demgegenüber jeweils ein Beurteilungszeitraum von 14 Tagen zugrunde. Damit ist weiterer Forschungsbedarf verbunden, wobei überprüft werden müsste, inwieweit sich unterschiedliche Zeitfenster auf die Reliabilität und Validität der Einschätzungen auswirken. Damit haben die hier berichteten Ergebnisse nur bedingten Anspruch auf Vergleichbarkeit zu den empirischen Befunden der anderen Studien.

#### *11.1.7 Zusammenfassung und Schlussfolgerungen*

**Zusammenfassend** kann festgehalten werden, dass der HoNOS-D ein praktikables Instrument darstellt und somit in der Routine Verwendung finden kann. Dies ist im Hinblick auf einen Einsatz innerhalb eines Klassifikationssystems von großer Bedeutung. Die Ergebnisse zeigen aber auch auf, dass zu einer Routineerhebung ohne Schulung zumindest eine Überarbeitung der Beurteilerinstruktionen und des Glossars nötig wären. Weiterhin kann festgehalten werden, dass die weitestgehende Unabhängigkeit der einzelnen Items bzw. Skalen der HoNOS-D nachgewiesen werden konnte. Die empirischen Befunde zeigen aber auch, dass einige Items keine unmittelbare Relevanz zur differenzierten Erfassung der Schwere einer psychischen Störung bei Patienten in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung aufweisen. Auch die empirischen Befunde zur Reliabilität (Interrater-Reliabilität und Test-Retest-Reliabilität) sind heterogen. Während für die Test-Retest-Reliabilität aufgrund der Höhe der erzielten Koeffizienten von einer weitestgehend reliablen Messung ausgegangen werden kann, erweisen sich die Einzelitems in der Analyse zweier ergänzend erhobener Teilstichproben hinsichtlich der Interraterreliabilität als defizitär. Dabei ist jedoch sicherlich auch der Aspekt bedeutsam, dass es sich bei den Ratern um nicht-trainierte Beurteiler gehandelt hat. Eine Überprüfung der Interraterreliabilitäten der Einzelitems der HoNOS-D *nach einer Schulung* der Beurteiler steht noch aus. Eine abschließende Beurteilung der empirischen Befunde zur generellen Eignung der HoNOS-D soll in einer Zusammenschau mit den empirischen Befunden zur spezifischen Eignung der HoNOS-D innerhalb der Fallgruppenbildung in Abschnitt 11.3 vorliegender Arbeit erfolgen.

**Insgesamt betrachtet**, kann aufgrund der empirischen Befunde festgehalten werden, dass der **Gesamtscore der HoNOS-D** in allen untersuchten Merkmalen hinsichtlich Praktikabilität

und Reliabilität befriedigende Resultate erzielte. Aufgrund dieser empirischen Befunde und aufgrund inhaltlicher Überlegungen ging er in den weiteren Analysen als Indikator zur differenzierten Erfassung der Schwere einer psychischen Störung in die Fallgruppenbildung von Patienten in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung ein.

## 11.2 Empirische Überprüfung der spezifischen Eignung der HoNOS-D innerhalb der Fallgruppenbildung

Die Entwicklung eines Klassifikationssystems in stationärer Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen erfolgte durch die theorie- und empiriegeleitete Ableitung von relevanten Prädiktoren des Ressourcenverbrauchs. In die Analyse gingen insgesamt 37 Kriterien ein, u.a. auch der „Gesamtscore der HoNOS-D“. Mittels Regression-Tree-Verfahren wurden anschließend 12 Prädiktoren zur Vorhersage der Behandlungsdauer extrahiert, welche 21 ressourcenhomogene Fallgruppen bildeten. Der so entstandene Entscheidungsbaum wies eine Gesamtvarianzaufklärung von 34.4% auf.

Bevor nun auf die Stellung des Gesamtscores der HoNOS-D und seine spezifische Eignung innerhalb des Fallgruppensystems eingegangen wird, soll im Folgenden zunächst eine **generelle Bewertung des resultierenden Klassifikationssystems** für Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung erfolgen.

### *11.2.1 Generelle Bewertung des Klassifikationssystems für Patienten mit psychischen Störungen*

Die **generelle Bewertung eines Klassifikationssystems** sollte verschiedene Punkte berücksichtigen. Es sollte die Güte der eingesetzten Instrumente beleuchtet werden und die Anzahl und klinische Plausibilität der ressourcenhomogenen Fallgruppen sowie die erzielte Gesamtvarianzaufklärung des Klassifikationsbaumes kritisch betrachtet werden.

Wie bereits erwähnt, sollte ein Klassifikationssystem zur Bildung von Fallgruppen verschiedene Anforderungen hinsichtlich der **Güte psychometrischer Eigenschaften** erfüllen und insbesondere in der Routine verwendbar sein. Betrachtet man die Ergebnisse, die im Metho-

denteil der vorliegenden Untersuchung berichtet werden, so kann festgehalten werden, dass die psychometrischen Eigenschaften der eingesetzten standardisierten und gut etablierten Instrumente hinsichtlich Praktikabilität und Reliabilität als weitestgehend erfüllt angesehen werden können. Die Ergebnisse der deskriptiven Analysen fehlender Werte zeigen befriedigende Resultate: So weisen nur zwei Skalen ca. 5% Missings auf (ADS-K, IIP-Subskala). Dennoch kann bereits an dieser Stelle festgehalten werden, dass das Erhebungsinstrument in seiner jetzigen Version für den Einsatz in der Routine noch nicht als ausreichend praktikabel (im Sinne einer kurzen Beantwortungszeit) gelten kann. Hinsichtlich der faktoriellen Validität konnte die postulierte Skalenstruktur in der konfirmatorischen Faktorenanalyse für 4 der 5 eingesetzten Selbst- und Fremdbeurteilungsverfahren bestätigt werden. Für die SF-8 Fremdeinschätzung zeigten sich zwar in der konfirmatorischen Faktorenanalyse keine befriedigenden Resultate, dafür ergaben sich aber zufrieden stellende interne Konsistenzen, so dass dieses Instrument auch mit hinreichender faktorieller Validität in die Fallgruppenanalysen einging. Die standardisierten Einzelitems, sowohl auf metrischem als auch auf nicht-metrischem Skalenniveau, wiesen heterogene Verteilungseigenschaften auf. Da es sich aber gemäß der Literatur um potentiell ressourcenverbrauchsrelevante Prädiktoren handelte, wurden sie trotz dieser Einschränkung in die Analyse mit aufgenommen.

**12 Prädiktoren** trugen signifikant zur Vorhersage der Behandlungsdauer bei und bildeten 21 ressourcenhomogene Fallgruppen mit einer Gesamtvarianzaufklärung von 34.4%. Im Einzelnen waren das: die Erstdiagnose zu Beginn der Behandlung, der Schweregradzusatz der Erstdiagnose, das Alter der Patienten, die Chronifizierung der Haupterkrankung, Beeinträchtigung in der gesundheitsbezogenen Lebensqualität (SF-8 Psychische Summenskala selbst und fremd beurteilt), **Belastung im Schweregrad (HoNOS-D)**, Belastung durch interpersonelle Probleme (IIP-25), Vorhandensein von Komorbidität, psychotherapeutische Vorerfahrung, aktuelle Partnersituation und der BMI der Patienten.

Zunächst fällt auf, dass die „Erstdiagnose“ als einer der zentralen Prädiktoren der Behandlungsdauer extrahiert wurde. Entgegen empirischer Befunde, wonach die **Diagnose** einen eher geringen Beitrag zur Varianzaufklärung des Ressourcenverbrauchs von Patienten mit psychischen Störungen leistet (2-8%, Jagdfeld et al., 2002; Mestel et al., 2002; Tritt et al., 2002), klärt dieses Merkmal in der vorliegenden Untersuchung einen wesentlich höheren Varianzanteil an der Behandlungsdauer auf. Das könnte eventuell damit zusammenhängen, dass in der Stichprobe ein besonders hoher Anteil von Patienten mit einer Essstörung vertreten war (ca.

15%, vgl. hierzu Tabelle 16 im Abschnitt 9.4.2), weswegen es eventuell zu einer Konfundierung von Erstdiagnose Essstörung und dem speziellen Behandlungskonzept dieser Klinik gekommen sein könnte. Diese Vermutung wird zusätzlich dadurch untermauert, dass diejenigen Patienten mit einer Essstörung sich auch in der ressourcenintensivsten Fallgruppe befanden. Dieses wiederum könnte dazu geführt haben, dass dadurch mehr Varianz an der Behandlungsdauer gebunden wurde und dieses zu einer Erhöhung der Varianzaufklärung der Diagnose geführt haben könnte.

In diesem Zusammenhang zeigen die Ergebnisse des Entscheidungsbaumes weiterhin, dass nicht allein die Diagnose zur Varianzaufklärung der Behandlungsdauer von Patienten mit psychischen Störungen in stationärer Rehabilitationsbehandlung beiträgt. Andere zentrale Prädiktoren beziehen sich vor allem auf Kriterien, die unter dem Schweregrad einer Erkrankung subsumiert werden können aber als auch auf verschiedene weitere patientenbezogene Merkmale (wie z.B. Probleme im interpersonellen Bereich). Bezogen auf die **Schwere einer Erkrankung** können somit die Befunde aus der Literatur bestätigt werden: Je höher die Belastung (durch Behandler eingeschätzt oder durch den Patienten selbst wahrgenommen), desto höher auch der durchschnittliche Ressourcenverbrauch (Taube et al., 1984b; Mitchell et al., 1987; Ashcraft et al., 1989; Horn et al., 1989; Bassler et al., 1995; Zielke et al., 1997; Buckingham et al., 1998; Borgart & Meermann, 1999; Kächele & für die Studiengruppe MZ-ESS, 1999; Spießl et al., 2002; Roick et al., 2002a).

Ein weiteres Resultat dieser Analyse ist die **Bildung von 21 ressourcenhomogenen Fallgruppen**. Dabei waren die **Patienten in der Gruppe mit den durchschnittlich längsten Behandlungsdauern** durch eine Essstörung als Erstdiagnose, die von den Behandlern als schwer eingestuft wurde, durch stärkere Belastung in der fremd eingeschätzten gesundheitsbezogenen Lebensqualität (Psychische Summenskala der SF-8-F) und durch eine höhere Anzahl ambulanter psychotherapeutischer Kontakte vor der Aufnahme gekennzeichnet. Diese Ergebnisse stehen im Einklang mit der Literatur und einer eigenen empirischen Untersuchung (Andreas et al., 2004a). Auch in der Studie von Andreas et al. (2004a) wiesen die Patienten der ressourcenintensivsten Fallgruppe u.a. eine Essstörung als Erstdiagnose auf. Interessant ist in diesem Zusammenhang, dass die Patienten der Gruppe mit den durchschnittlich längsten Behandlungsdauern auch schon vor der Aufnahme mehr psychotherapeutische, ambulante Behandlungstermine in Anspruch genommen haben. Dieses Ergebnis steht im Einklang mit

den empirischen Befunden von Borgart und Meermann (1999), die aufzeigen konnten, dass mit steigender Anzahl an psychotherapeutischen/ambulanten Vorbehandlungen auch die späteren Behandlungszeiten der Patienten anstieg. Damit lässt sich auch vermuten, dass das Merkmal psychotherapeutische Vorerfahrung eine Proxyvariable der Schwere der Erkrankung darstellen könnte. Im Gegensatz dazu sind die **Patienten mit den durchschnittlich kürzesten Behandlungsdauern** durch eine Neurotische-, Belastungs- oder somatoforme Störung als Erstdiagnose, durch eine Erkrankungsdauer unter 3 Jahren, durch geringere Probleme im interpersonellen Bereich und durch einen niedrigeren BMI (unter 23) beschreibbar. Auch dieses Ergebnis steht im Einklang mit der Literatur und mit den empirischen Ergebnissen einer eigenen Untersuchung (Andreas et al., 2004a). Patienten, die weniger Ressourcen verbrauchen bzw. beanspruchen, sind tendenziell auch eher weniger hinsichtlich krankheitsrelevanter Merkmale belastet (wie z.B. Schweregrad, Chronifizierung).

Hinsichtlich der **Anzahl der extrahierten Fallgruppen** lässt sich zeigen, dass diese in einem befriedigenden, unter ökonomischen Gesichtspunkten handhabbaren Bereich liegt und mit den berichteten Fallgruppenzahlen anderer Studien übereinstimmt (Gaines et al., 2003; Taube et al., 1984b; Mitchell et al., 1987, vgl. hierzu Tabelle 2 im Abschnitt 3.2).

Das **Entscheidungsbaummodell der Gesamtstichprobe** erreicht eine **Varianzaufklärung** von **34.4%**. Der Kennwert liegt im Vergleich zu den empirischen Befunden der internationalen, komplementär zu den DRG-Systemen entwickelten Klassifikationsmodelle in einem akzeptablen Bereich und geht zudem deutlich über den in einer eigenen Untersuchung zur Vorhersage des Ressourcenverbrauchs ermittelten Betrag von 17.5% hinaus (Andreas et al., 2004a). Dieser Unterschied, wie bereits oben erwähnt, könnte aber auch wieder durch eine Konfundierung zwischen Erstdiagnose Essstörung und spezifischem Behandlungskonzept zu erklären sein. Bei der Diskussion der Höhe dieses Kennwertes im Vergleich zu allgemeinen DRG-Systemen, die bei Einschluss von Extremwerten Varianzaufklärungen der Behandlungsdauer von ca. 47% erzielen (Fischer, 2002), sollte außerdem berücksichtigt werden, dass diese DRG-Systeme die bereits erbrachten therapeutischen Leistungen am Ende der Behandlung mit in die Analyse einspeisen. Es handelt sich somit um teilweise retrospektive Systeme. Eine Zielsetzung der vorliegenden Untersuchung war es jedoch, zu überprüfen, ob und, wenn ja, welche Patientenmerkmale die Behandlungsdauer **vorhersagen**.

Ein weiteres Problem, dass bei der Diskussion der Höhe der zu erzielenden Varianzaufklärung zum Tragen kommt, besteht in der relativen Invarianz der Kriteriumsvariable Behandlungsdauer (Variationskoeffizient der Behandlungsdauer = 36%). Vermutlich könnte bei einer Flexibilisierung der Behandlungsdauern auch eine Erhöhung der Varianzaufklärung erreicht werden. Daraus kann geschlussfolgert werden, dass eine Anpassung der Verweildauer nötig wäre, um eine höhere Varianzaufklärung eines Klassifikationssystems von Patienten mit psychischen Störungen zu erreichen und damit ein adäquates Modell vorhalten zu können. Im Zuge einer derartigen Flexibilisierung von Behandlungszeiten wäre jedoch in jedem Falle der (katamnestische) Behandlungserfolg zu überprüfen.

Abschließend sollen bereits an dieser Stelle einige **methodische Probleme und Beschränkungen** der vorliegenden Untersuchung zur Entwicklung eines Klassifikationssystems in stationärer Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen diskutiert werden. Eine zusammenfassende Einordnung der empirischen Befunde soll allerdings unter dem Gesichtspunkt der Generalisierbarkeit in Abschnitt 11.3 erfolgen. Eines der viel diskutierten Probleme der Regression-Tree-Analyse liegt in der **Variabilität der extrahierten Baumstrukturen** (Schweizer & Vach, 1999). Da dieses Verfahren bei der Selektion der Prädiktoren auf den Vergleichen zwischen $chi^2$- und F-Werten der einzelnen Prädiktoren beruht, kann es bereits bei geringsten Modifikationen (z.B. Eliminieren von Ausreißern) zu anderen Baumstrukturen mit einer anderen Auswahl von Prädiktoren kommen.

In **zwei Analysen** wurde in der vorliegenden Arbeit die **Variabilität der extrahierten Baumstruktur** und **der Fallgruppen** untersucht. Dabei wurde an 10 bzw. an 5 zufällig gezogenen Stichproben überprüft, ob sich die gerade beschriebenen 12 Prädiktoren mittels Regression-Tree-Verfahren erneut extrahieren lassen bzw. die gerade aufgeführten 21 ressourcenhomogenen Fallgruppen mittels Linearer Regressionsanalyse hinsichtlich ihrer Rangfolge in der Behandlungsdauer bestätigen lassen. Die Ausführungen zum methodischen Vorgehen finden sich in Abschnitt 10.3.1.2, Kapitel 10 vorliegender Arbeit.

Hinsichtlich der Stabilität der Baumstruktur als Ganzes (vgl. Gesamtbaum Abbildung 44, in Abschnitt 10.3.1) zeigte sich in zehn Zufallsstichproben, dass die ersten drei Baumebenen (oder Verästelungsebenen) eine moderate bis noch zufrieden stellende Stabilität innerhalb der Zufallsstichproben aufwiesen (60% bis 100% Stabilität). Die unterste oder vierte Ebene des Gesamtbaumes wies nur eine moderate Stabilität auf (0% bis 60%). Auch hinsichtlich der

Stabilität der 21 ressourcenhomogenen Fallgruppen erwiesen sich im Mittel 14 Fallgruppen bezüglich ihrer Behandlungsdauerreihung in fünf Zufallsstichproben als stabil, dagegen zeigte nur jeweils 1 Fallgruppe starke Abweichungen. In beiden Analysen zur Überprüfung der Stabilität des Gesamtbaumes bzw. der ressourcenhomogenen Fallgruppen konnten Varianzaufklärungskennwerte erzielt werden, die mit einer Spannweite von ca. 30% bis 37% in einem zufrieden stellenden Bereich um die Gesamtvarianzaufklärung des Hauptbaumes von 34.4% lagen. Aus diesen Ergebnissen kann damit an dieser Stelle bereits festgehalten werden, dass sich sowohl der Gesamtbaum als auch die extrahierten Fallgruppen nur als bedingt stabil erwiesen. Dieser Befund macht allerdings auch deutlich, dass noch zu überprüfen bleibt, ob größere Stichprobenumfänge mit einer höheren Repräsentanz von Patienten verschiedener Diagnosegruppen und unterschiedlichster Fachkliniken, auch zu stabileren Prädiktoren und damit auch zu stabileren Fallgruppen führen. Denn es wäre auch denkbar, dass je heterogener die Stichprobe wird, die Wahrscheinlichkeit für eine Instabilität zunimmt.

Konform zur Hypothese 2a erwies sich, wie oben bereits genannt, der **Gesamtscore der HoNOS-D innerhalb der Fallgruppenbildung** als ein relevanter Prädiktor zur Vorhersage der Behandlungsdauer. Die Diskussion der differenzierten Analysen und Ergebnisse zur spezifischen Eignung der HoNOS-D sollen nun im Folgenden vor dem Hintergrund der Hypothesen 2a bis 2c vorgenommen werden.

### *11.2.2 Zur Bedeutung der HoNOS-D bei der Fallgruppenbildung*

In der Hypothese 2a war postuliert worden, dass der „Gesamtscore der HoNOS-D“ mittels Regression-Tree-Verfahren extrahiert wird und somit einen Prädiktor innerhalb der Fallgruppenbildung von Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung darstellt.

Aufgrund der in Abschnitt 10.3.1 bereits berichteten empirischen Ergebnisse kann die Hypothese 2a angenommen werden. Der Gesamtscore der HoNOS-D wurde mittels Regression-Tree-Verfahren extrahiert und stellt ein Prädiktor innerhalb der Fallgruppenbildung dar. Wie die Ergebnisse weiterhin aufzeigen, ist er insbesondere ein Indikator zur Vorhersage des Ressourcenverbrauchs bei Patienten mit affektiven Störungen, die darüber hinaus älter als 29 Jahre sind. Damit erweist er sich als ein zentrales Merkmal zur Vorhersage des Ressourcenverbrauchs bei einer relevanten Behandlungsgruppe von Patienten in stationärer psychosoma-

tischer/psychotherapeutischer Rehabilitationsbehandlung. So stellen die affektiven Störungen in diesem Versorgungsbereich die Hauptbehandlungsgruppe dar (in der vorliegenden Stichprobe bilden sie mit 38% die häufigste Diagnosegruppe, vgl. hierzu auch, Schulz et al., in Druck), und die Patienten in der stationären psychosomatischen/psychotherapeutischen Rehabilitationsbehandlung sind zudem im Durchschnitt eher älter (M = 42 Jahre, SD = 14 Jahre, vgl. Tabelle 16). Fasst man nun noch einmal beide Kriterien (affektive Störungen und älter als 29 Jahre) zusammen, so lässt sich festhalten, dass sich der Gesamtscore der HoNOS-D bei 35% der Patienten der vorliegenden Untersuchungsstichprobe als bedeutsam erweist.

Ein weiteres zentrales Ergebnis vorliegender Analyse ist, dass der Gesamtscore der HoNOS-D in Kombination mit drei weiteren Merkmalen (Partnersituation, Psychische Summenskala der SF-8-S und Komorbidität) insgesamt **6 ressourcenhomogene Fallgruppen** bildet. Verglichen mit anderen Prädiktoren des Gesamtbaumes, z.B. der Psychischen Summenskala der SF-8-F (vgl. Abbildung 44, in Abschnitt 10.3.1), nimmt der Gesamtscore der HoNOS-D damit den differenziertesten Anteil an der Bildung der Fallgruppen ein, was als zusätzlicher Hinweis auf seine Relevanz zur Vorhersage der Behandlungsdauer gewertet werden kann.

In den differenzierten Analysen zur **Stabilität des Gesamtbaumes** und der **extrahierten Fallgruppen** wurde der „Gesamtscore der HoNOS-D" in 6 der 10 Zufallsstichproben als Prädiktor des Ressourcenverbrauchs extrahiert. Hinsichtlich der Rangfolge in der Behandlungsdauer erwiesen sich im Mittel 4 von den insgesamt 6 HoNOS-D bezogenen, ressourcenhomogenen Fallgruppen als identisch. Aus diesen Befunden lässt sich wiederum ableiten, dass sich die Fallgruppen, die u.a. durch den Gesamtscore der HoNOS-D gebildet werden, als bedingt stabil erwiesen. Dieses Resultat könnte damit einhergehen, dass der Gesamtscore der HoNOS-D insgesamt betrachtet nur einen marginalen Anteil an der Gesamtvarianzaufklärung des Fallgruppenbaumes leistet, weswegen er dann auch in Zufallsstichproben nicht gänzlich stabil repliziert werden kann. Dieser Aspekt soll nun aber im Vergleich zu einem anderen, annähernd äquivalenten Fremdeinschätzungsscore vor dem Hintergrund der Hypothesen 2b und 2c ausführlicher diskutiert werden. Da sich diese beiden Hypothesen auf den Vergleich des Gesamtscores der HoNOS-D zu einem äquivalenten Fremdeinschätzungsscore beziehen, sollen sie im Folgenden auch zusammen erörtert werden.

### *11.2.3 Zur Varianzaufklärung des Gesamtscores der HoNOS-D und zur Bedeutung innerhalb eines „Prädiktoren-Kern-Sets“*

Bevor die empirischen Befunde zur Bedeutung des Gesamtscores der HoNOS-D im Vergleich zu einem anderen Fremdeinschätzungsinstrument diskutiert werden, soll kurz eine **generelle Bewertung** der empirischen Ergebnisse der Linearen Regressionsanalyse im Hinblick auf den Gesamtbaum erfolgen.

Zunächst lässt sich wieder generell festhalten, dass sich die **12 Prädiktoren** des Gesamtbaumes auch in der Linearen Regressionsanalyse hinsichtlich ihrer Bedeutung zur Varianzaufklärung der Behandlungsdauer befriedigend replizieren lassen. So ist z.B. das Merkmal Erstdiagnose einer der zentralen Prädiktoren im Gesamtbaum und klärt auch in der Linearen Regressionsanalyse zusammengenommen den größten Anteil an der Varianz der Behandlungsdauer auf. Auch der insgesamt aufgeklärte Varianzanteil der Linearen Regressionsanalyse entspricht der durch das Regression-Tree-Verfahren ermittelten Gesamtvarianz von 34.4%.

In Hypothese 2b war postuliert worden, dass sich in einer Multiplen Regressionsanalyse mit den 12 Prädiktoren des Gesamtbaumes für den Gesamtscore der HoNOS-D im Vergleich zu einem äquivalenten Fremdeinschätzungsscore (Psychische Summenskala der SF-8-F) höhere signifikant positive Regressionsgewichte aufzeigen lassen. In der Linearen Regressionsanalyse ergab sich für den Vergleich zwischen dem Gesamtscore der HoNOS-D und der Psychischen Summenskala der SF-8-F, dass beide Prädiktoren signifikant positive Regressionsgewichte aufwiesen. Die Hypothese 2b hat sich demzufolge nur zum Teil bestätigen lassen.

In Hypothese 2c war erwartet worden, dass der Gesamtscore der HoNOS-D bei einem „Prädiktoren-Kern-Set“[21] mittels Regression-Tree-Verfahren extrahiert wird und einen höheren

[21] Das **„Prädiktoren-Kern-Set“** wurde unter zwei Prämissen gebildet. Zum ersten basiert es auf den 12 Prädiktoren, die sich bereits im Regression-Tree-Verfahren als signifikant erwiesen haben (vgl. Abbildung 44) und zum zweiten wurden diejenigen der 12 Prädiktoren ausgewählt, die bereits in der Routineerhebung von psychosomatischen/psychotherapeutischen Fachkliniken im Rahmen der internen Qualitätssicherung vorhanden sein könnten. Es wurden folgende Merkmale als „Prädiktoren-Kern-Set“ definiert: Erstdiagnose, Alter, Partnersituation, ambulante psychotherapeutische Vorerfahrung, Komorbidität und Chronifizierung der Erkrankung. Darüber hinaus wurde in der ersten Analyse der Prädiktor Gesamtscore der HoNOS-“ und in der zweiten Analyse der Prädiktor Psychische Summenskala der SF-8-F in das Regression-Tree-Verfahren eingespeist.

Beitrag zur Varianzaufklärung liefert als ein vergleichbarer Fremdeinschätzungsscore (die Psychische Summenskala der SF-8-F). Dabei zeigte sich, dass sowohl der Gesamtscore der HoNOS-D als auch die Psychische Summenskala der SF-8-F im Kontext des „Prädiktoren-Kern-Sets“ mittels Regression-Tree-Verfahren extrahiert wurden. Somit kann auch die Hypothese 2c nur zum Teil bestätigt werden.

Bei Vergleich der Ergebnisse dieser beiden Regression-Tree-Analysen (siehe auch Abbildung 46 und Abbildung 47) zeigt sich für den Gesamtscore der HoNOS-D eine leichte Überlegenheit gegenüber der Psychischen Summenskala der SF-8-F. So bildet der Gesamtscore der HoNOS-D eine ressourcenhomogene Fallgruppe mehr, womit eine tendenziell höhere Differenziertheit deutlich wird. Auch hinsichtlich der Position in den beiden Bäumen zeigt sich, dass der Gesamtscore der HoNOS-D bereits auf der dritten Verästelungsebene extrahiert wird, die Psychische Summenskala der SF-8-F dagegen erst auf der vierten Verästelungsebene. Beide Prädiktoren spielen bei der Hauptbehandlungsgruppe der Affektiven Störungen eine zentrale Rolle, wobei der Gesamtscore der HoNOS-D hierbei eine bedeutsamere Stellung einnimmt. Es zeigt sich aber auch, dass die Psychische Summenskala der SF-8-F zusätzlich noch in der Diagnosegruppe der Essstörungen von Bedeutung ist, wobei es sich hierbei um eine Störungsgruppe handelt, die eher weniger im Fokus der stationären Rehabilitationsbehandlung steht. Diese ist bei Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung nur in ca. 3% der Fälle vertreten (Verband Deutscher Rentenversicherungsträger, 2004, zitiert nach Dirmaier, 2005).

**Zusammenfassend** lässt sich festhalten, dass keine klare Überlegenheit des Gesamtscores der HoNOS-D im Vergleich zur Psychischen Summenskala der SF-8-F festgestellt werden konnte. Dafür können nun verschiedene Gründe in Betracht kommen. Zum einen konnte gezeigt werden, dass beide Prädiktoren nur einen sehr geringen Anteil an der Gesamtvarianzaufklärung der Behandlungsdauer liefern und der Hauptvarianzanteil auf das Merkmal Erstdiagnose fällt. Dieser Umstand könnte eine weitere Aufschlüsselung zur Differenzierung zwischen beiden Prädiktoren erschwert haben. Ein weiterer Punkt betrifft die Schulung der Rater. Der Beurteilung im HoNOS-D lag zwar ein differenziertes Glossar zugrunde, allerdings bleibt unklar, in wie weit dieses von den Beurteilern genutzt wurde und in wie weit Verständnisprobleme auftraten. Beim SF-8-F, welcher in dieser Form erstmalig eingesetzt wurde, lag kein Glossar zugrunde, sondern die Beurteiler schätzten die relevanten Items anhand der vorge-

benden Ankerpunkte ein (vgl. Anhang 16.2.1). Aufgrund der befriedigenden Befunde zu den psychometrischen Kennwerten der Psychischen Summenskala der SF-8-F (vgl. hierzu Tabelle 20 in Abschnitt 9.5.2.2.2) kann dabei auch geschlussfolgert werden, dass im SF-8-F anscheinend keine Schulung notwendig ist, was eher für den Einsatz der SF-8-F spricht. Hinsichtlich dieser Annahme stehen aber noch Überprüfungen weiterer psychometrischer Eigenschaften der SF-8-F aus.

Auf der Grundlage dieser empirischen Befunde sind nun differenzierte Empfehlungen zu einem Einsatz der HoNOS-D innerhalb eines Klassifikationssystems für Patienten mit psychischen Störungen ableitbar. Diese Aspekte sollen abschließend in Abschnitt 11.5 erörtert werden.

#### *11.2.4 Zur Bedeutung von Selbst- vs. Fremdeinschätzungsinstrumenten innerhalb der Fallgruppenbildung*

In internationalen, komplementär zu allgemeinen DRG-Systemen entwickelten Klassifikationssystemen für Patienten mit psychischen Störungen in stationärer psychiatrischer/psychotherapeutischer Behandlung werden fast ausschließlich Fremdeinschätzungsinstrumente zur Fallgruppenbildung verwendet. Deswegen interessierte in einer explorativen Nebenfragestellung im Vergleich dazu die **Bedeutsamkeit von Selbsteinschätzungsinstrumenten** innerhalb der Fallgruppenbildung. Diesem Interesse liegen zum einen ökonomische Überlegungen zugrunde, zum anderen könnte durch die Nutzung von Selbsteinschätzungsinstrumenten dem so genannten „Gaining“ von Therapeuten (Hirdes et al., 2003) entgegengewirkt werden: „For example, variables that cannot be observed by an independent, external rater would be easy to manipulate in the hope of gaining additional funding. The greater the reliance on subjective interpretation by the assessor, the greater the risk that the variable can be systematically overestimated with relative impunity.” (zitiert na. Hirdes et al., 2003, S. 8).

Im Rahmen des „Prädiktoren-Kern-Sets“ wurde nun überprüft, in wie weit sich eine zu den beiden Fremdeinschätzungsscores (Gesamtscore der HoNOS-D und Psychische Summenskala der SF-8-F) vergleichbare Selbsteinschätzungsskala (Psychische Summenskala der SF-8-S) als relevanter Prädiktor zur Vorhersage der Behandlungsdauer nachweisen lässt. Dabei spielte auch eine Rolle, welchen Varianzaufklärungsanteil dieser Prädiktor erzielen konnte.

Die Ergebnisse des Regression-Tree-Verfahrens zeigen, dass die Psychische Summenskala der SF-8-S bei zwei Diagnosegruppen von Patienten mit psychischen Störungen (Patienten mit einer Affektiven Störung und einer Neurotischen-, somatoformen oder Belastungsstörung extrahiert wird und ihre Position im Entscheidungsbaum mit der Stellung des Gesamtscores der HoNOS-D vergleichbar ist. Das Modell mit dem Prädiktor Psychische Summenskala der SF-8-S weist im Vergleich zu den beiden fremd beurteilten Variablen dagegen eine tendenziell weniger differenzierte Fallgruppenbildung auf (13 ressourcenhomogene Fallgruppen) und erreicht eine etwas niedrigere Varianzaufklärung an der Behandlungsdauer (31.53%). Auch in der Multiplen Regressionsanalyse liegt das standardisierte ß-Gewicht der Psychischen Summenskala der SF-8-S unterhalb der Gewichte der beiden fremd eingeschätzten Scores. Die Unterschiede werden allerdings nicht signifikant.

Daraus kann die Schlussfolgerung gezogen werden, dass auch Selbsteinschätzungsinstrumente ausreichende Berücksichtigung innerhalb der Fallgruppenbildung für Patienten mit psychischen Störungen in diesem Versorgungsbereich finden sollten. Diskutiert man allerdings die gewonnenen empirischen Befunde und das entwickelten Klassifikationssystem im Hinblick auf eine Anwendung auch in anderen Versorgungsbereichen, z.B. der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie, so sollten dort zumindest ergänzend Fremdeinschätzungsinstrumente zum Einsatz kommen, da sich in diesen Versorgungsbereichen häufiger Patienten befinden, die zum Zeitpunkt ihrer Aufnahme nicht in der Lage sind, über sich selbst Auskunft geben zu können (z.B. Patienten mit einer akuten schizophrenen Episode).

### *11.2.5 Methodische Beschränkungen und Probleme*

Ein zentraler Aspekt bei der differenzierten Bildung von Fallgruppen betrifft die methodische Voraussetzung, dass genügend **Varianz in der Kriteriumsvariablen Behandlungsdauer** vorhanden sein sollte (Buckingham et al., 1998). Diese war in der vorliegenden Analyse nur zum Teil erfüllt. Der ermittelte Variationskoeffizient (VK) der Behandlungsdauer zeigte mit 36.7% eine gerade noch ausreichende Variation auf. Auch das Ergebnis der ermittelten Varianzaufklärung des Gesamtbaumes weist auf eine diesbezügliche Einschränkung hin. Die erzielte Varianzaufklärung von 34.4% kann vor dem Hintergrund empirischer Befunde zwar als zufrieden stellend beurteilt werden (Richter, 2001), es werden jedoch auch ca. 65% der Varianz der Behandlungsdauer nicht erklärt. In der Literatur konnte aufgezeigt werden, dass oft

weitere strukturelle Merkmale der Einrichtungen (wie z.B. die Lage der Klinik oder differenzielle Behandlungsangebote) zu einer substantiellen Erhöhung der aufgeklärten Varianz führen können (Richter, 2001). Dementsprechend sollte an dieser Stelle bei einer Einführung von Fallpauschalen für den Bereich der stationären Behandlung von Patienten mit psychischen Störungen empfohlen werden, zunächst eine Vergleichbarkeit der Einrichtungen herzustellen und eine Adjustierung unterschiedlicher Risikostrukturen vorzunehmen (Richter, 2001). Eine besonders gewichtige Einschränkung stellt die Begrenzung der Behandlungsdauer und des Ressourceneinsatzes durch Vorgaben der Kostenträger dar. Hier wäre es nötig, unter Sicherstellung des katamnestischen Outcomes der Patienten, eine Flexibilisierung der Behandlungsdauer zu ermöglichen und dieses zu überprüfen.

Eine weitere methodische Beschränkung der vorliegenden Untersuchung bezieht sich auf das **Problem der Variabilität der Baumstrukturen**. Es konnte gezeigt werden, dass die extrahierten Prädiktoren sich nur zum Teil in anderen Zufallsstichproben als stabil erwiesen. Das gibt Hinweise darauf, dass eine weitere Überprüfung der gewonnenen Prädiktoren an anderen, noch zu erhebenden Stichproben notwendig ist.

## 11.3 Generalisierbarkeit der Befunde

Zur Beurteilung der **Generalisierbarkeit** der empirischen Befunde vorliegender Arbeit und des entwickelten Klassifikationssystems für Patienten mit psychischen Störungen gilt zunächst zu prüfen, in wie weit die erhobene Stichprobe als repräsentativ für das Gesamtpatientenklientel der 11 verschiedenen Rehabilitationsfachkliniken gelten kann. Zudem gilt es zu überprüfen, in wie weit sich die vorliegende Stichprobe als repräsentativ für den Versorgungsbereich der stationären psychosomatischen/psychotherapeutischen Rehabilitationsbehandlung erweist. Darüber hinaus kann dann die Übertragbarkeit der empirischen Befunde bzw. des entwickelten Klassifikationssystems auf andere Versorgungsbereiche der Behandlung psychisch Kranker untersucht werden.

Die **Drop-Out-Patienten** (Nonresponder und Partial-Nonresponder) waren zu Beginn der stationären Behandlung nach Einschätzung der Therapeuten weniger motiviert und symptomatisch stärker belastet. Die Partial-Nonresponder (als eine Teilgruppe der Drop-Out-Patienten) waren weniger häufig berufstätig und hatten häufiger einen Rentenantrag gestellt als die Patienten der Untersuchungsstichprobe. Da die Unterschiede zwischen den Drop-Out-

Patienten und den Studienteilnehmern jedoch größtenteils nur mit kleiner Effektstärke bestanden, kann für die vorliegende Untersuchungsstichprobe eine weitestgehende Repräsentativität für das Gesamtklientel dieser 11 verschiedenen Rehabilitationsfachkliniken angenommen werden.

Hinsichtlich der **Repräsentativität** der Patienten der Untersuchungsstichprobe **für die stationäre psychosomatische/psychotherapeutische Rehabilitationsbehandlung** lässt sich zunächst festhalten, dass die Patienten vorliegender Untersuchungsstichprobe zu 74% Frauen und im Durchschnitt 42 Jahre alt waren. Als häufigste Erstdiagnose wiesen sie eine Affektive Störung (38%) auf, gefolgt von Patienten mit einer Essstörung (15%), einer Belastungs- und Anpassungsstörung (13%) oder einer Angststörung (9%). Im Vergleich dazu sind die Patienten mit psychischen/psychosomatischen Erkrankungen in stationärer Rehabilitationsbehandlung im Durchschnitt etwas jünger (ca. 40 Jahre alt). Auch werden durchschnittlich weniger Frauen in den stationären Einrichtungen der psychosomatischen/psychotherapeutischen Rehabilitation behandelt (66%). Die Patienten in diesen Einrichtungen leiden im Mittel ebenso häufig unter depressiven Störungen (37%), seltener jedoch unter Essstörungen (3%) und Anpassungsstörungen (7%) und häufiger unter Angststörungen (11%) (Verband Deutscher Rentenversicherungsträger, 2004, zitiert nach Dirmaier, 2005). In diesem Zusammenhang fällt insbesondere der deutlich höhere Anteil der Patienten einer Spezialklinik für Essstörungen ins Gewicht, weswegen für die vorliegende Untersuchungsstichprobe nur eine eingeschränkte Repräsentativität für das Gesamtklientel der Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung verzeichnet werden kann.

Vor dem Hintergrund einer potentiellen **DRG-Einführung im akut-stationären Versorgungsbereich** soll im Folgenden insbesondere die Überprüfung einer Übertragbarkeit der Untersuchungsergebnisse auf diesen Bereich im Fokus stehen. Die Notwendigkeit zur Prüfung auf Repräsentativität für den ambulanten Versorgungssektor psychisch Kranker erscheint zum jetzigen Zeitpunkt nicht angezeigt.

Wie Schulz und Koch (2002) in ihrer Expertise zur stationären Versorgung psychisch Kranker in Norddeutschland aufzeigten, besteht ein hoher Überschneidungsbereich zwischen den stationären Versorgungsbereichen von Patienten mit psychischen Störungen. Die empirische Datenbasis ist allerdings derzeit noch als mangelhaft zu bewerten. Zielsetzung vorliegender

Arbeit war es u.a. ein Klassifikationssystem für den Bereich von Patienten mit psychischen Störungen in stationärer Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen zu konstruieren. In diesem Zusammenhang wurde insbesondere die spezifische Eignung der HoNOS-D zur Fallgruppenbildung untersucht. Bei der HoNOS handelt es sich um ein Instrument, welches primär zur Beurteilung schwerstbelasteter Patienten in stationärer psychiatrischer/psychotherapeutischer Behandlung konstruiert wurde (Wing et al., 1998). Eine Überprüfung der psychometrischen Eigenschaften in diesem Bereich steht für die HoNOS-D allerdings noch aus und erscheint vor dem Hintergrund der Testkonstruktion bzw. der hier ermittelten Verteilungseigenschaften besonders wünschenswert. Weiterhin kann geschlussfolgert werden, dass die extrahierten Prädiktoren nur geringe rehabilitationsspezifische Anteile besitzen und durchaus auch im Bereich der stationären psychiatrischen/psychotherapeutischen Versorgung in Zusammenhang zum Ressourcenverbrauch stehen könnten. Dabei ist allerdings einschränkend hinzuzufügen, dass bestimmte Diagnosegruppen (z.B. Patienten mit einer psychotischen Erkrankung) häufiger in der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie vertreten sind (Schulz et al., in Druck) und damit wiederum andere Prädiktoren bei der Vorhersage der Behandlungsdauer an Bedeutung gewinnen könnten. Deswegen erscheint eine Überprüfung der extrahierten Prädiktoren und damit auch der HoNOS-D innerhalb der Fallgruppenbildung für Patienten mit psychischen Störungen auch in der Krankenhausbehandlung im Bereich Psychiatrie und Psychotherapie sowie Psychosomatik und Psychotherapie möglich und wünschenswert.

**Zusammenfassend** lässt sich vor dem Hintergrund dieser Überlegungen aussagen, dass eine Generalisierbarkeit der empirischen Befunde vorliegender Untersuchung auf den Gesamtbereich der Rehabilitationsbehandlung von Patienten mit psychischen/psychosomatischen Störungen und auf den Versorgungsbereich der Krankenhausbehandlung im Bereich Psychosomatik und Psychotherapie sowie Psychiatrie und Psychotherapie eher nicht möglich erscheint. Dazu trägt auch der Umstand bei, dass in der vorliegenden Untersuchungsstichprobe ein hoher Varianzanteil an der Behandlungsdauer auf Patienten mit einer Essstörung entfiel.

**Weitergehende Schlussfolgerungen**

Wie gerade aufgezeigt werden konnte, handelt es sich bei der Untersuchungsstichprobe um eine weitestgehend repräsentative Stichprobe von Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung. Der Ho-

NOS-D erwies sich als ein praktikables und weitgehend reliables Instrument mit Schwächen in der Interraterreliabilität. Insbesondere zeigte sich der Gesamtscore der HoNOS-D hinsichtlich der geprüften psychometrischen Eigenschaften den Einzelitems überlegen. Auch die spezifische Eignung der HoNOS-D innerhalb der Fallgruppenbildung konnte nachgewiesen werden. Allerdings zeigte sich keine Überlegenheit gegenüber einem vergleichbaren alternativen Fremdeinschätzungsscore. Diese Befunde sollen nun die Ausgangsbasis für weitere Überlegungen darstellen.

Im Folgenden sollen nun Empfehlungen zur Überarbeitung der HoNOS-D und Empfehlungen zum Einsatz des Gesamtscores der HoNOS-D zur Fallgruppenbildung (in Abwägung zum Einsatz eines anderen Fremdeinschätzungsscores) gegeben werden.

## 11.4 Empfehlungen zur Überarbeitung der HoNOS-D

Entsprechend den bereits aufgezeigten Schwächen sollen folgende Veränderungen vorgeschlagen werden:

**a) Überarbeitung der Beurteilerinstruktionen:**

Um eine Anwendung der HoNOS-D auch in der Routine zu gewährleisten, erscheint es wünschenswert, eine prägnante Beurteilerinstruktion zu erarbeiten und den Beurteilern zur Verfügung zu stellen. Dabei wird davon ausgegangen, dass eine Schulung aus ökonomischen Gründen nicht realisierbar erscheint.

Folgende Punkte können hierbei von besonderer Relevanz sein:

- Kurze Erläuterung des Prinzips der Beurteilung im HoNOS-D (z.B. Sicherstellung der Unabhängigkeit der Items)
- Information darüber, dass die Vergabe der Kategorie 9 „nicht bekannt“ möglichst vermieden werden sollte, da sie als fehlender Wert gezählt werden muss

- Information darüber, dass bei der Einschätzung von Item 2 „Selbstverletzung“ nicht nur selbstverletzendes Verhalten, sondern auch Selbstmordgedanken bzw. –gefährdung mit einzubeziehen sind

- Hinweis für die Beurteiler, dass die Ankerpunkte im Glossar *als Beispiele* fungieren

- Hinweis bei Item 8 „Andere Symptome“ jeweils das Problem auszuwählen, worin der Patient *außerdem*, also unabhängig von den bereits bewerteten Problemen, am meisten belastet ist

- Prägnante Erläuterungen der Items 10 „Probleme mit alltäglichen Aktivitäten“, 11 „Probleme durch die Wohnbedingungen“ und 12 „Probleme durch die Bedingungen im Beruf und im Alltag“

Es ist auch denkbar, dass die Beurteiler eine Fallvignette mit entsprechender Beurteilung im HoNOS-D zur Verfügung gestellt bekommen. Damit könnte wiederum eine Erhöhung der Vergleichbarkeit der Einschätzungen erzielt werden. Eine bereits erarbeitete Beurteilerinstruktion befindet sich im Anhang 16.6 vorliegender Arbeit.

**b) Modifikation des Items 8 „Andere psychische und verhaltensbezogene Probleme“ der HoNOS-D**

Als besonders relevant für den Bereich der Behandlung von Patienten mit psychischen/psychosomatischen Erkrankungen erwies sich das Item 8 „Andere Symptome“. Gleichzeitig wurden in diesem Item aufgrund der durchgeführten empirischen Analysen noch die meisten Schwachstellen festgestellt.

Folgende Modifikationen sind denkbar:

- Differenzierte Beurteilung für jedes der 10 verschiedenen aufgeführten Probleme

- Erarbeitung von detaillierten Ankerpunkten in den einzelnen Ausprägungen (0 bis 4) der zehn verschiedenen Probleme

*Zum Beispiel für die Kategorie „Phobisch“ des Items 8 und einem Rating von 3 „eher schweres Problem“*

*3 Erhebliches Problem, wenn der Patient z.B. das spezifische phobische Objekt vermeidet (z.B. keine Benutzung des Fahrstuhls bei Claustrophobie mehr möglich)*

- Um eine internationale Vergleichbarkeit zu gewährleisten, sollte aber auch weiterhin eine Auswahl des am schwersten eingeschätzten der zehn Probleme erfolgen

Ein Vorschlag zur Modifikation des Items 8 der HoNOS-D befindet sich im Anhang 16.6. Die differenzierten Ankerpunkte für das Glossar müssten allerdings noch erarbeitet werden.

**c) Überarbeitung des Glossars der HoNOS-D**

Auch das Glossar weist bei genauerer Betrachtung partielle Schwächen auf. So sind einige Ankerpunkte nicht mit differenzierten Beispielen für Patienten aus allen Versorgungsbereichen psychisch Kranker versehen. Deswegen wäre hier denkbar:

- Differenziertere Ausgestaltung der Beispiele für Patienten mit psychischen Störungen aus allen Versorgungsbereichen

- Präzisierung der Überschriften der Einzelitems (z.B. Item 2)

- Ausgabe der 1-seitigen Beurteilerinstruktion als Ergänzung zum Glossar

Abschließend sollte noch angemerkt werden, dass die Überarbeitungen keine gravierenden Veränderungen beinhalten sollten, um noch eine Vergleichbarkeit der empirischen Ergebnisse nationaler Studien zu Resultaten anderer Untersuchungen zu gewährleisten.

## 11.5 Empfehlungen zum Einsatz der HoNOS-D innerhalb eines Fallgruppensystems

Auch im Hinblick auf die spezifische Eignung der HoNOS-D innerhalb der Fallgruppenbildung sollen abschließend noch einige Empfehlungen gegeben werden. Diese sollen vor allem vor dem Hintergrund des Vergleiches mit einem äquivalenten Fremdeinschätzungsscore, der Psychischen Summenskala der SF-8-F, vorgenommen werden.

Vergleicht man die beiden Prädiktoren Gesamtscore der HoNOS-D und Psychische Summenskala der SF-8-F miteinander, so ist der Einsatz innerhalb der Fallgruppenbildung mit Vor- und Nachteilen verbunden. Eine Bewertung soll daher unter verschiedenen Gesichtspunkten vorgenommen werden.

*Ökonomische Gesichtspunkte*

Bei der SF-8-F handelt es sich um ein fremdeingeschätztes Instrument zur Erfassung gesundheitsbezogener Lebensqualität, welches über acht Items einen Psychischen und einen Somatischen Summenscore erfasst. Bei der HoNOS-D handelt es sich um Fremdeinschätzungsinstrument zur differenzierten Erfassung der Schwere einer psychischen Störung. Es beinhaltet 12 Items, die zu einem Gesamtscore zusammengefasst werden können. Außerdem sollte der Einschätzung im HoNOS-D vermutlich eher eine Schulung zugrunde liegen als einer Einschätzung im SF-8-F.

Daraus kann abgeleitet werden, dass unter ökonomischen Gesichtspunkten (8 Items der SF-8-F vs. 12 Items der HoNOS-D; Schulung vs. keine Schulung) der Einsatz der SF-8 als Fremdeinschätzungsinstrument der HoNOS-D überlegen ist.

*Gesichtspunkt klinischer Relevanz*

Beim differenzierten Vergleich beider Prädiktoren innerhalb der Fallgruppenbildung zeigte sich, dass der Gesamtscore der HoNOS-D sowohl eine zentralere Position im Fallgruppenbaum einnahm als auch bei einer bedeutsamen Hauptbehandlungsgruppe extrahiert wurde. Unter dem Gesichtspunkt der klinischen Relevanz zeigte sich hierbei der Gesamtscore der HoNOS-D der Psychischen Summenskala der SF-8-F überlegen.

*Gesichtspunkt internationaler Vergleichbarkeit und Generalisierbarkeit auf andere Versorgungsbereiche*

Mit der HoNOS-D liegt ein Fremdeinschätzungsinstrument vor, welches international weit verbreitet ist. Damit können die empirischen Ergebnisse auch auf breiter internationaler Ebene diskutiert werden. Auch der SF-8 ist ein weit verbreitetes und international eingesetztes Instrument, jedoch nicht in der hier verwendeten Fremdeinschätzungsversion, sondern bislang

nur als Selbsteinschätzungsinstrument. Im Hinblick auf eine internationale Vergleichbarkeit vorliegender Befunde zeigt sich somit der HoNOS-D als überlegen.

Hinsichtlich einer Anwendbarkeit in anderen Versorgungsbereichen von Patienten mit psychischen Störungen scheinen beide Instrumente geeignet zu sein. Beim HoNOS-D wäre dabei auch noch zu überprüfen, ob sich die Verteilungseigenschaften einiger Items (z.B. Item 6 „Halluzinationen und Wahnvorstellungen") noch verbessern lassen. Damit ist noch weiterer Forschungsbedarf verbunden, der im Folgenden abschließend näher skizziert werden soll.

### 11.6 Weiterer Forschungsbedarf

Wie in den vorangegangenen Kapiteln gezeigt wurde, erwies sich der HoNOS-D als ein praktikables Fremdeinschätzungsinstrument zur differenzierten Erfassung der Schwere einer psychischen Störung. Hinsichtlich der Reliabilität bestehen allerdings partiell noch Defizite. In dem Zusammenhang ist an dieser Stelle hervorzuheben, dass eine Überprüfung der Interraterreliabilitäten nach einer Schulung der Beurteiler noch aussteht. Im Rahmen der Vorbereitung solcher Schulungen wäre auch die Erarbeitung einer modifizierten und differenzierteren Version des Glossars der HoNOS-D wünschenswert.

Im Hinblick auf den routinemäßigen Einsatz der HoNOS-D ist weiterer Forschungsbedarf mit der modifizierten Version der Beurteilerinstruktion verbunden. Hier wäre im Rahmen von kontrollierten, experimentellen Designs der Frage nachzugehen, welche Effekte eine Schulung im Vergleich zur Gabe der modifizierten Beurteilerinstruktion und eines modifizierten Glossars aufweisen würde.

Schließlich wäre im Rahmen der hier vorgeschlagenen Überarbeitungen der HoNOS-D, insbesondere des Items 8 „Andere psychische und verhaltensbezogene Probleme", insgesamt eine vertiefende Überprüfung psychometrischer Eigenschaften vorzunehmen. Dabei sollte vor allem das modifizierte Item 8 im Rahmen der Erhebung von Beurteilerübereinstimmungen hinsichtlich seiner Interraterreliabilität untersucht werden.

Ein weiterer Aspekt der Überprüfung könnte sich auf eine differenzierte Untersuchung der Validität der HoNOS-D im Vergleich zu einem anderen gut etablierten Fremdeinschätzungsinstrument, z.B. dem Beeinträchtigungs-Schwere-Score beziehen. In diesem Zusammenhang

könnte auch der Frage nachgegangen werden, in wie weit sich der HoNOS-D als ein differenzielles Zuweisungs- und Steuerungsinstrument in der Behandlung von Patienten mit psychischen Störungen erweist.

Wie die vorausgegangenen Kapitel weiter aufzeigten, hat sich der HoNOS-D innerhalb des entwickelten Klassifikationssystems von Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung als ein Prädiktor zur Vorhersage der Behandlungsdauer bewährt. Hier wäre in einem nächsten Schritt zu überprüfen, in wie weit sich der Gesamtscore der HoNOS-D in anderen Stichproben als Prädiktor replizieren lässt. Dieser Entwicklungsschritt könnte auch mit einer Reduktion des insgesamt noch nicht als praktikabel zu bewertenden Klassifikationssystems einhergehen. Dazu wäre es auch nötig, diejenigen Prädiktoren aus dem nun vorliegenden Klassifikationssystem zu extrahieren, die den größten differenziellen Beitrag an der Varianzaufklärung der Behandlungsdauer leisten, und ´den Variablensatz im Hinblick auf die Konstruktion eines „Prädiktoren-Kern-Sets“ zu reduzieren. In diesem „Prädiktoren-Kern-Set“ sollten dann jene Kriterien enthalten sein, die sich sowohl hinsichtlich der Fallgruppenbildung als relevant erwiesen haben, als auch unmittelbar in der Routineversorgung verfügbar sind. Ein erster Versuch zur Bildung eines solchen „Prädiktoren-Kern-Sets“ wurde in vorliegender Arbeit bereits realisiert. Die Überprüfung dieses Sets an anderen Stichproben, auch aus den unterschiedlichsten Versorgungsbereichen psychisch Kranker, wäre mit weiteren empirischen Studien verbunden.

## 12 Zusammenfassung

Die differenzierte Erfassung des Schweregrades von Patienten mit psychischen Störungen nimmt gerade in den letzten Jahren unter ökonomischen Gesichtspunkten eine zentrale Bedeutung ein. Insbesondere innerhalb von international entwickelten Patientenklassifikationssystemen, die über allgemeine Diagnosis-Related-Groups-Systeme hinausgehen (DRGs), stellt die Erfassung der Schwere der Erkrankung eine zentrale Dimension dar, um den Ressourcenverbrauch in stationärer psychotherapeutischer Behandlung adäquater abbilden zu können. Dabei zeigt sich vor allem die Relevanz eines Fremdeinschätzungsinstrumentes, der „Health of the Nation Outcome Scales (HoNOS)“. Zielsetzung vorliegender Arbeit ist es, in einem ersten Schritt die deutsche Version der Health of the Nation Outcome Scales, HoNOS-D als ein Fremdeinschätzungsinstrument zur differenzierten Erfassung der Schwere einer psychischen Störung hinsichtlich ihrer psychometrischen Eigenschaften, insbesondere der Praktikabilität und Reliabilität zu untersuchen. Ein weiteres Ziel besteht darin, die spezifische Eignung der HoNOS-D innerhalb der Entwicklung eines Klassifikationssystems für Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung zu prüfen.

Das Fremdeinschätzungsinstrument HoNOS wurde zunächst in einem mehrstufigen, experten-gestützten Konsensusprozess von der Arbeitsgruppe für Psychotherapie- und Versorgungsforschung in die deutsche Sprache übersetzt. Die empirische Überprüfung der generellen und spezifischen Eignung der HoNOS-D erfolgte anhand einer weitestgehend repräsentativen, konsekutiv erhobenen Stichprobe von N = 1.918 Patienten aus 11 verschiedenen Rehabilitationsfachkliniken des Indikationsbereiches „Psychische Störungen“. An zwei ergänzend erhobenen Stichproben aus dem Bereich der stationären psychosomatischen/psychotherapeutischen Krankenhausbehandlung und dem Bereich der Rehabilitationsbehandlung Suchtkranker wurden außerdem die Interraterreliabilitäten der HoNOS-D untersucht.

Der HoNOS-D weist im Rahmen der empirischen Überprüfung psychometrischer Eigenschaften eine zufrieden stellende Praktikabilität und mit Einschränkungen auch Reliabilität auf. Die empirischen Ergebnisse hinsichtlich der Interraterreliabilität können als unbefriedigend bewertet werden. Es zeigte sich, dass der Gesamtscore der HoNOS-D den Einzelitems in den

psychometrischen Überprüfungen überlegen war, weswegen er als Indikator der Schwere einer psychischen Störung in die Fallgruppenbildung einging.

Im Rahmen der Entwicklung eines Klassifikationssystems für Patienten mit psychischen Störungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung erwies sich der Gesamtscore der HoNOS-D als ein Prädiktor zur Vorhersage der Behandlungsdauer. Im Vergleich zu einem alternativ eingesetzten Fremdeinschätzungsscore konnte er allerdings keine Überlegenheit demonstrieren.

Auf der Grundlage der durchgeführten empirischen Analysen werden Empfehlungen zur Modifikation der HoNOS-D und zum Einsatz des Instrumentes innerhalb eines Patientenklassifikationssystems für Patienten mit psychischen Störungen gegeben.

## 13 Literaturverzeichnis

Adams, M., Palmer, A., O'Brien, J. T. & Crook, W. (2000). Health of the Nation Outcome Scales for Psychiatry: Are they valid? *Journal of Mental Health, 9*(2), 193-198.

Alden, L. E., Wiggins, J. S. & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment, 55*, 521-536.

Amelang, M. & Zielinski, W. (1997). *Psychologische Diagnostik und Intervention*. Berlin: Springer-Verlag.

American Psychiatric Association. (1996). *Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Dt. Bearbeitung und Einführung von Henning Saß*. Göttingen: Hogrefe.

American Psychiatric Association. (2002). GAF: Skala zur Globalen Erfassung des Funktionsniveaus. In E. Brähler & B. Strauß (Hrsg.), *Diagnostische Verfahren in der Psychotherapie* (S. 154-158). Göttingen: Hogrefe-Verlag.

Andreas, S., Dirmaier, J., Koch, U. & Schulz, H. (2005). *Fallgruppen in der stationären Rehabilitation von Patienten mit psychischen/psychosomatischen Erkrankungen* (unveröffentlichter Abschlussbericht des Projektes C2 im NVRF). Hamburg: Institut und Poliklinik für Medizinische Psychologie, Universitätsklinikum Hamburg-Eppendorf.

Andreas, S., Dirmaier, J., Lang, K., Watzke, B., Koch, U., Ranneberg, J. & Schulz, H. (2004a). Lassen sich Zusammenhänge zwischen Patientenmerkmalen und dem fachspezifischen Ressourcenverbrauch in der stationären Behandlung von Patienten mit psychischen/psychosomatischen Störungen bestimmen? *Psychotherapie Psychosomatik Medizinische Psychologie, 54*, 1-9.

Andreas, S., Harfst, T., Dirmaier, J., Kawski, S., Koch, U. & Schulz, H. (2004b). A psychometric evaluation of the German version of the "Health of the Nation Outcome Scales, HoNOS-D": On the feasibility and reliability of a clinician-rated measure of severity in patients with mental disorders. *Eingereicht zur Publikation in Psychopathology*.

Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMDP). (1996). AMDP-System. In C. I. P. S. (CIPS) (Hrsg.), *Internationale Skalen für Psychiatrie* (Band 4. überarb. und erw. Aufl.). Göttingen: Beltz-Test.

Arbeitskreis OPD. (2004). *Operationalisierte Psychodynamische Diagnostik - OPD. Grundlagen und Manual*. Bern: Verlag Hans Huber.

Asendorpf, J. & Wallbott, H. G. (1979). Maße der Beobachterübereinstimmung: Ein systematischer Vergleich. *Zeitschrift für Sozialpsychologie, 10*, 243-252.

Ashcraft, M. L., Fries, B. E., Nerenz, D. R., Falcon, S. P., Srivastava, S. V., Lee, C. Z., Berki, S. E. & Errera, P. (1989). A Psychiatric Patient Classification System: An alternative to Diagnosis-Related Groups. *Medical Care, 27*(5), 543-557.

Audin, K., Marginson, F. R., Clark, J. M. & Barkham, M. (2001). Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services. *British Journal of Psychiatry, 178*, 561-566.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2000). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Berlin: Springer.

Barth, M. (1997). Konzeption und Evaluation multipler Regressionsanalysen in der anwendungsorientierten klinisch-psychologischen Forschung. In B. Strauß & J. Bengel (Hrsg.), *Forschungsmethoden in der Medizinischen Psychologie. Jahrbuch der Medizinischen Psychologie 14* (S. 146-160). Göttingen: Hogrefe.

Bassler, M., Krauthauser, H. & Hoffmann, S. O. (1995). Welche Faktoren beeinflussen die Dauer von stationärer Psychotherapie? *Psychotherapie Psychosomatik Medizinische Psychologie, 45*, 167-175.

Baumann, U. & Stieglitz, R.-D. (1989). Evaluation des AMDP-Systems anhand der neueren Literatur (1983-1987) - Überblicksarbeit. *Fortschritte Neurologie Psychiatrie, 57*, 357-373.

Beckmann, D. & Richter, H. E. (1983). *Gießen-Test*. Bern: Huber.

Biggs, D., De Ville, B. & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics, 18*(1), 49-62.

Borgart, E.-J. & Meermann, R. (1999). Bedingungsfaktoren unterschiedlicher Behandlungsdauer bei Angststörungen im Rahmen stationärer Verhaltenstherapie. *Psychotherapie Psychosomatik Medizinische Psychologie, 49*, 109-113.

Bortz, J. (1995). *Statistik für Sozialwissenschaftler* (2. Aufl. Aufl.). Berlin: Springer.

Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2. Aufl.). Berlin: Springer.

Brähler, E. & Scheer, J. (1983). *Der Gießener Beschwerdebogen: GBB*. Bern: Huber.

Brähler, E. & Scheer, J. (2002). GBB: Der Gießener Beschwerdebogen. In E. Brähler, J. Schumacher & B. Strauß (Hrsg.), *Diagnostische Verfahren in der Psychotherapie* (S. 159-163). Göttingen: Hogrefe-Verlag.

Brooks, R. (2000). The reliability and validity of the Health of the Nation Outcome Scales: Validation in relation to patient derived measures. *Australian and New Zealand Journal of Psychiatry, 34*, 504-511.

Browne, S., Doran, M. & McGauran, S. (2000). Health of the Nation Outcome Scales (HoNOS): Use in an Irish psychiatric outpatient population. *Irish Journal of Psychological Medicine, 17*(1), 17-19.

Buckingham, B., Burgess, P., Solomon, S., Pirkis, J. & Eagar, K. (1998). *Developing a Casemix Classification for Mental Health Services: Summary*. Canberra: Commonwealth Department of Health & Family Services.

Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.

Bullinger, M. & Kirchberger, I. (1998). *SF-36. Fragebogen zum Gesundheitszustand*. Göttingen: Hogrefe.

Bundesministerium für Gesundheit. (2002). *Informationen zum Gesetz zur Einführung des diagnose-orientierten Fallpauschalensystems für Krankenhäuser (Fallpauschalengesetz, FPG)*.http://www.bm.gesundheit.de/fpg.doc.

Carthew, R., Elphick, M. & Page, R. (2003). *Report on the development of Mental Health Groupings*. Winchester.

Clade, H. (2000). Fallpauschalensystem: Kopernikanische Wende? *Deutsches Ärzteblatt, 97*(28-29), 1939-1941.

Coffey, R. M. (1999). Casemix information in the United States: Fifteen years of management and clinical experience. *CASEMIX Quarterly, 1*(1), 1-15.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale, NJ: Lawrence Erlbaum.

Collegium Internationale Psychiatriae Scalarum (CIPS). (1996a). Brief Psychiatric Rating Scale (BPRS). In Collegium Internationale Psychiatriae Scalarum (CIPS) (Hrsg.), *Internationale Skalen für Psychiatrie* (Band 4. überarb. und erw. Aufl., S. 105-111). Göttingen: Beltz-Test.

Collegium Internationale Psychiatriae Scalarum (CIPS). (1996b). CGI: Clinical Global Impressions. In Collegium Internationale Psychiatriae Scalarum (CIPS) (Hrsg.), *Internationale Skalen für Psychiatrie* (Band 4. überarb. und erw. Aufl., S. 147-151). Göttingen: Beltz-Test.

Collegium Internationale Psychiatriae Scalarum (CIPS). (1996c). Das AMDP-System: Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie. In Collegium Internationale Psychiatriae Scalarum (CIPS) (Hrsg.), *Internationale Skalen für Psychiatrie* (Band 4. überarb. und erw. Aufl., S. 225-233). Göttingen: Beltz-Test.

Commonwealth Department of Health and Family Services. (1998a). *Development of the Australian Refined Diagnosis Related Groups (AR-DRG) classification version 4 Volume 1: Summary of changes for the AR-DRG classification version 4.0.* Canberra.

Commonwealth Department of Health and Family Services. (1998b). *Development of the Australian Refined Diagnosis Related Groups (AR-DRG) Classification Version 4: Volume 1 Summary of changes for the AR-DRG Classification Version 4.0*. Canberra: Commonwealth of Australia.

Derogatis, L. R. (Hrsg.). (1977). *SCL-90-R, administration, scoring & procedures manual-I for the R(evised) version.* Baltimore: John Hopkins University School Medicine.

Deutsche Gesellschaft für Psychotherapeutische Medizin (DGPM). (2001). *Die Einführung der DRG´s in der Psychosomatik und Psychotherapie.* Vortrag gehalten auf 52. DKPM-Arbeitstagung, Bad Honnef.

Dilling, H., Mombour, W. & Schmidt, M. H. (1993). *Internationale Klassifikation psychischer Störungen: ICD-10 Kapitel V (F) - Klinisch-diagnostische Leitlinien*. Bern: Verlag Hans Huber.

Dirmaier, J. (2005). *Profil Psychotherapeutischer Zielsetzungen: Konstruktion und psychometrische Überprüfung eines Instruments zur therapeutenseitigen Messung von Therapiezielen.* Unveröffentlichte Dissertation Type, Universität Hamburg, Hamburg.

Dirschedl, P. (1992). Klassifikationsbäume - Grundlagen und Neuerungen. In W. Fleischer, M. Nagel & R. Ostermann (Hrsg.), *Interaktive Datenanalyse mit ISP* (S. 15-30). Essen: Westarp Wissenschaften.

Egner, U. & Verbarg, A. (2001). Das DRG-System im Krankenhaus und seine Auswirkungen auf die Rehabilitation. *Deutsche Angestellten Versicherung, 11*, 1-6.

Elsner, H., Bätz, B., Magerkurth, T. & Rüth, S. (2003). German Diagnosis Related Groups: Problemfelder im Bereich psychischer Störungen am Beispiel der Krankenhausbehandlung Alkoholabhängiger. *Der Nervenarzt*, 1-14.

Endicott, J., Spitzer, R. L., Fleiss, J. L. & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatry disturbance. *Archives of General Psychiatry, 33*, 766-771.

English, J. T., Sharfstein, S., Scherl, D. J., Astrachan, B. & Muszynski, I. L. (1986). Diagnosis-Related Groups and general hospital psychiatry: The APA study. *American Journal of Psychiatry, 143*(2), 131-139.

Fetter, R. B., Shin, Y., Freeman, J. L., Averill, R. F. & Thompson, J. D. (1980). Case mix definition by diagnosis-related groups. *Medical Care, 18*, 1-53.

Fischer, W. (1997). *Patientenklassifikationssysteme zur Bildung von Behandlungsfallgruppen im stationären Bereich: Prinzipien und Beispiele*. Bern: Bundesamt für Sozialversicherung.

Fischer, W. (2000). *Zur Kostenhomogenität in verschiedenen DRG-Systemen*. Z/I/M. http://www.fischer-zim.ch/.

Fischer, W. (2001). *Das australische AR-DRG-System als Beispiel einer ärztlich-ökonomischen Klassifikation*. Z/I/M. http://www.fischer-zim.ch/artikel/ARDRG-0105-SGMI.htm

Fischer, W. (2002). *Kurzgutachten: "Diagnosis Related Groups (DRG´s) und verwandte Patientenklassifikationssysteme", Version 1.2*. http://www.fischer-zim.ch

Fischer, W. (2003). *Die Bedeutung von Pflegediagnosen in Gesundheitsökonomie und Gesundheitsstatistik*. http://www.fischer-zim.ch/artikel/Pflege-Diagnosen-9901-WEG-htm

Franke, G. (1995). *SCL-90-R. Die Symptom-Checkliste von Derogatis - Deutsche Version*. Göttingen: Beltz.

Frick, U., Barta, W. & Binder, H. (2001). Fallpauschalen in der stationär-psychiatrischen Versorgung: Empirische Evaluation im Land Salzburg. *Psychiatrische Praxis, 28*, S55-S62.

Fritze, J., Saß, H. & Möller, H.-J. (1999). GKV-Gesundheitsreform 2000 - Kabinettsentwurf des Gesetzes: Gemeinsame Stellungnahme der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) und der Arbeitsgemeinschaft für Neuropsychopharmakologie und Pharmakopsychiatrie (AGNP). *Nervenarzt, 70*, 766-771.

Gaines, P., Bower, A., Buckingham, W., Eager, K., Burgess, P. & Green, J. (2003). New Zealand Mental Health Classification and Outcomes Study: Final report. Auckland: Health Research Council of New Zealand.

Garms-Homolová, V., von Kardoff, E., Fuchs, H., Lüngen, M. & Lauterbach, K. (2002). *Zur Entwicklung von Patientenklassifikationssystemen für die Medizinische Rehabilitation* (unveröffentlichtes Positionspapier). Berlin: Institut für Gesundheitsanalysen und soziale Konzepte e.V.

Gerdelmann, W. (1997). Vergütungssysteme in der Krankenhauspsychiatrie der Zukunft. In K. D. Hildemann & P. Potthoff (Hrsg.), *Psychotherapie - Quo vadis?: Ziele, Effektivität und Kosten in Psychiatrie und Psychosomatik* (S. 55-64). Göttingen: Hogrefe.

Gerdes, N. & Jäckel, W. H. (1992). Indikatoren des Reha-Status (IRES) - Ein Patientenfragebogen zur Beurteilung von Rehabilitationsbedürftigkeit und -erfolg. *Die Rehabilitation, 31*, 73-79.

Goldney, R. D., Fisher, L. J. & Walmsley, S. H. (1998). The Health of the Nation Outcome Scales in psychiatric hospitalisation: A multicentre study examining outcome and prediction of length of stay. *Australian and New Zealand Journal of Psychiatry, 32*, 199-205.

Goodman, S. H., Sewell, D. R., Cooley, E. L. & Leavitt, N. (1993). Assessing levels of adaptive functioning: The Role Functioning Scale. *Community Mental Health Journal, 29*, 119-131.

Haaf, H.-G. & Röckelein, E. (2002). *Einführung der DRG: Eine grundlegende Reform der Vergütung von Krankenhausleistungen mit weitreichenden Folgen.* www.dgvt.de/politik/berufsbild/ein_drg.htm.

Hamilton, A. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, Psychiatry, 23*, 56-62.

Hamilton, M. (2002). Die Hamilton Depression Skala (HAMD). In E. Brähler, J. Schumacher & B. Strauß (Hrsg.), *Diagnostische Verfahren in der Psychotherapie* (S. 183-186). Göttingen: Hogrefe Verlag.

Harfst, T., Andreas, S., Dirmaier, J., Kawski, S., Koch, U. & Schulz, H. (2004a, 8th-11th September 2004). *The role of the German version of "The Health of the Nation Outcome Scales, HoNOS-D" in measuring symptom severity II: Analysis of validity.* Vortrag gehalten auf 34th EABCT Annual Congress, Manchester.

Harfst, T., Dirmaier, J., Butsch, R., Kurtz von Aschoff, C., Koch, U. & Schulz, H. (2001). Klinische Validität der Subskalen der SCL-90-R: Die Übereinstimmung von klinisch-diagnostischen Kategorien, Fremdeinschätzungen und Patientenselbstbeurteilungen im Verlauf einer stationären psychosomatischen Rehabilitation. *DRV-Schriften, 26*, 100-101.

Harfst, T., Dirmaier, J., Koch, U. & Schulz, H. (2004b). Entwicklung und psychometrische Überprüfung einer multidimensionalen Kurzform des Inventars zur Erfassung interpersonaler Probleme. *DRV-Schriften, 52*, 53-55.

Harfst, T., Koch, U., Kurtz von Aschoff, C., Nutzinger, D. O., Rüddel, H. & Schulz, H. (2002). Entwicklung und Validierung einer Kurzform der Symptom Checklist-90-R. *DRV-Schriften, 33*, 71-73.

Häuser, W., Zimmer, C., Wilhelm, R., Klein, W. & Krause-Wichmann, D. (2004). International psychosomatic medicine within the German Diagnosis Related Groups System. *Psycho-Social-Medicine, 1*, 1-8.

Hautzinger, M. & Bailer, M. (1993). *Allgemeine Depressionsskala (ADS). Die deutsche Version des CES-D.* Weinheim: Beltz.

Hautzinger, M. & Bailer, M. (2002). Die Allgemeine Depressions Skala (ADS). In E. Brähler, J. Schumacher & B. Strauß (Hrsg.), *Diagnostische Verfahren in der Psychotherapie* (S. 25-28). Göttingen: Hogrefe Verlag.

Hautzinger, M., Bailer, M., Worrall, H. & Keller, F. (1995). *Das Beck Depressionsinventar (BDI).* Bern: Huber.

Herrmann, C. H., Buss, U. & Snaith, R. P. (1995). *HADS-D - Hosptial Anxiety and Depression Scale - Deutsche Version: Ein Fragebogen zur Erfassung von Angst und Depressivität in der somatischen Medizin.* Bern: Verlag Hans Huber.

Heuft, G. & Senf, W. (1998). *Praxis der Qualitätssicherung in der Psychotherapie. Das Manual der Psy-BaDo*. Stuttgart: Thieme.

Hiller, W., Zerssen, D. v., Mombour, W. & Wittchen, H.-U. (1986). *Inpatient Multidimensional Psychiatric Scale. Deutsche Bearbeitung*. Weinheim: Beltz.

Hirdes, J. P., Fries, B. E., Botz, C., Ensley, C., Marhaba, M. & Pérez, E. (2003). *The System for Classification of In-Patient Psychiatry (SCIPP): A new case-mix methodology for mental health*.Unveröffentlichtes Manuskript, Waterloo ON, Canada.

Hoff, P. (1997). *Historische Aspekte psychopathologischer Befunderhebung*. Göttingen: Hogrefe-Verlag.

Horgan, C. & Jencks, S. F. (1987). Research on psychiatric classification and payment systems. *Medical Care, 25*(9), S22-S36.

Horn, S. D., Chambers, A. F., Sharkey, D. P. & Horn, R. A. (1989). Psychiatric Severity of Illness: A Case Mix Study. *Medical Care, 27*(1), 69-84.

Horowitz, L. M., Strauß, B. & Kordy, H. (1994). *Inventar zur Erfassung interpersonaler Probleme - Deutsche Version - (IIP-D)*. Weinheim: Beltz.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Irps, S., Roeder, N. & Müller, M. L. (2002). *Informationen zu German Refined Diagnosis Related Groups*. http://drg.uni-muenster.de/de/informationen/infogdrgs.html.

Jacobson, N. S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.

Jagdfeld, F. H., von Wachter, M. & Petzold, E. R. (2002). Erste Erfahrungen mit DRGs an einer psychosomatischen Universitätsklinik (Abstract). *Psychotherapie Psychosomatik Medizinische Psychologie, 52*, 94.

Janiszewski, J. & Pfaff, M. (2001). Das DRG-Kalkulationsmodell. *f&w, 3*, 221-228.

Jerschke, S., Meixner, K., Richter, H. & Bohus, M. (1998). Zur Behandlungsgeschichte und Versorgungssituation von Patientinnen mit Borderline-Persönlichkeitsstörung in der Bundesrepublik Deutschland. *Fortschritte Neurologie Psychiatrie, 66*, 545-552.

Kächele, H. & für die Studiengruppe MZ-ESS. (1999). Eine multizentrische Studie zu Aufwand und Erfolg bei psychodynamischer Therapie von Essstörungen. *Psychotherapie Psychosomatik Medizinische Psychologie, 49*, 100-108.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119-127.

Kawski, S. (2004). *Ergebnisbericht zur Erhebung der Strukturqualität: Gesamtbericht der Pilotphase 2003/2004*. Hamburg: Institut für Medizinische Psychologie, UKE.

Kawski, S. & Koch, U. (2002). Zum Stand der Qualitätssicherung in der Rehabilitation: Zur Entwicklung der medizinischen Rehabilitation in den 90er-Jahren. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 45*, 260-266.

Keller, A. (2000). *Die Klassifikation psychischer Störungen nach DSM-IV mit Hilfe eines strukturierten diagnostischen Interviews (F-DIPS) – Eine Untersuchung der Retest-Reliabilität und der Validität.* Unveröffentlichte Dissertation Type, Dresden.

Kluge, H., Hülsmann, S., Kopf, A., Angermeyer, M. C. & Becker, T. (2002). Stationäre psychiatrische Behandlungsdauer: Eine statistische Analyse auf Grundlage einer Basisdokumentation. *Krankenhauspsychiatrie, 13*, 104-110.

Koch, U., Arentewicz, G., Tiefensee, J., Bergelt, C. & Bauer, C. (1998). *Reha-Qualitätssicherungsprogramm der Rentenversicherung: Programmpunkt 2 - Patiententherapiepläne* (Unveröffentlichter Abschlussbericht). Hamburg: Medizinische Psychologie, UKE.

Koch, U. & Potreck-Rose, F. (1994). Stationäre psychosomatische Rehabilitation - ein Versorgungssystem in der Diskussion. In B. Strauß & A. E. Meyer (Hrsg.), *Psychoanalytische Psychosomatik - Theorie, Forschung und Praxis* (S. 193-212). Stuttgart: Schattauer.

Krautgartner, M., Scherer, M. & Katschnig, H. (2002). Psychiatrische Krankenhaustage: Wer konsumiert die meisten? Eine Record-Linkage-Studie über fünf Jahre in einem österreichischen Bundesland. *Psychiatrische Praxis, 29*, 355-363.

Kruckenberg, P., Wolfersdorf, M., Bauer, M., Kunze, H., Fritze, J. & Schmauß, M. (2001). Vergütung psychiatrischer Leistungen im neuen Krankenhaus-Entgeltsystem (DRG-System): Stellungnahme gegenüber der Deutschen Krankenhausgesellschaft (DKG). *Nervenarzt, 11*, 894-896.

Kunze, H. & Kaltenbach, L. (1996). *Psychiatrie-Personalverordnung* (3. erw. Aufl.). Stuttgart: Kohlhammer.

Laaser, U., Röttger-Liepmann, B., Breckenkamp, J. & Bublitz, K. (2000). Der Einfluß von Fallpauschalen auf die stationäre Versorgung. *Pflege, 13*, 9-15.

Laireiter, A.-R., Baumann, U. & Stieglitz, R.-D. (2001). Soziodiagnostik: Soziales Netzwerk, Soziale Unterstützung und Soziale Anpassung. In R.-D. Stieglitz & H. J. Freyberger (Hrsg.), *Psychodiagnostik in Klinischer Psychologie, Psychiatrie und Psychotherapie* (Band 2. überarb. und erw. Auflage, S. 210-228). Stuttgart: Thieme-Verlag.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lang, K. (2003). *Behandlungsabbrüche und Therapiemotivation in der stationären Rehabilitation von Patienten mit psychischen Erkrankungen - Entwicklung und empirische Überprüfung eines Vorhersagemodells. Dissertation.* Unveröffentlichte Dissertation Type, Universität Hamburg.

Lange, C. & Heuft, G. (2002). Die Beeinträchtigungsschwere in der psychosomatischen und psychiatrischen Qualitätssicherung: Global Assessment of Functioning Scale (GAF) vs. Beeinträchtigungs-Schwere-Score (BSS). *Zeitschrift für Psychosomatische Medizin und Psychotherapie, 48*, 256-269.

Lauterbach, K. & Lüngen, M. (2000a). *DRG-Fallpauschalen: Eine Einführung. Anforderungen an die Adaption von Diagnosis-Related Groups in Deutschland. Gutachten im Auftrag des AOK-Bundesverbandes*. Stuttgart: Schattauer.

Lauterbach, K. & Lüngen, M. (2000b). Neues Entgeltsystem nach US-Muster. *Deutsches Ärzteblatt, 97*(8), 392-395.

Laux, L., Glanzmann, P., Schaffner, P. & Spielberger, C. D. (1981). *STAI - Das State-Trait-Angstinventar. Theoretische Grundlagen und Handanweisungen*. Weinheim: Beltz.

Lemon, S. C., Roy, J., Clark, M., Friedmann, P. D. & Rakowski, W. (2003). Classification and Regression Tree Analysis in Public Health: Methodological review and comparison with Logistic Regression. *Annals of Behavioral Medicine, 26*(3), 172-181.

Little, R. & Rubin, D. (2002). *Statistical Analysis with Missing Data*. New York: John Wiley.

Lorr, M., Jenkins, R. L. & Holsopple, J. G. (1953). Multidimensional scale for rating psychiatric patients: Hospital form. *Veterans Administration Technical Bulletin, 43*, 10-507.

Lorr, M., Klett, C. J., McNair, D. M. & Lasky, J. J. (1963). *Manual: Inpatient Multidimensional Psychiatric Scale*. Palo Alto, Cal.: Consulting Psychologists Press.

Magidson, J. (1997). The CHAID approach to segmentation modeling: Chi-squared automatic interaction detection. In R. P. Bagozzi (Hrsg.), *Advanced methods of marketing research* (S. 118-160). Malden, Massachusetts: Blackwell.

Margraf, J., Schneider, S., Soeder, U., Neumer, S. & Becker, E. (1996). *F-DIPS. Diagnostisches Interview bei psychischen Störungen (Forschungsversion).*Unveröffentlichtes Manuskript, TU Dresden.

Maylath, E. (2000). DRGs in der psychiatrischen Krankenhausfinanzierung am Beispiel Ungarns: Ein Modell für Deutschland? *Gesundheitswesen, 62*, 633-645.

McClelland, R., Trimble, P., Fox, M. L., Stevenson, M. R. S. & Bell, B. (2000). Validation of an outcome scale for use in adult psychiatric practice. *Quality in Health Care, 9*, 98-105.

Mestel, R., Vogler, J. & Klingelhöfer, J. (2002). Zusammenhänge zwischen Haupt- und Nebendiagnosen mit der Behandlungsdauer in der stationären Psychosomatik. *DRV-Schriften, 33*, 502-503.

Meu. (2001). DRG-Viele Aufgaben müssen noch erledigt werden: Das Eckpunktepapier des BMG zur Einführung des DRG-Fallpauschalensystems in der Diskussion. *f&w*, 212-217.

Mezzich, J. E. (1991). Architecture of clinical information and prediction of service utilization and cost. *Schizophrenia Bulletin, 17*, 469-474.

Mitchell, J. B., Dickey, B., Liptzin, B. & Sederer, L. J. (1987b). Bringing psychiatric patients into the Medicare prospective payment system: Alternatives to DRGs. *Am J Psychiatry, 144*(5), 610-615.

Musiol, G. (1999). *Zielgruppenbestimmung*. Hamburg: Verlag Dr. Kovac.

Musiol, G. & Steinkamp, G. (1998). *CHAID: Ein Instrument für die empirische Marktforschung*. Braunschweig: Vieweg Verlag.

National Institute of Mental Health. (1976). 028 CGI. Clinical Global Impressions. In W. Guy & R. R. Bonato (Hrsg.), *Manual for the EDCEU Assessment Battery* (Band 2. rev. Version). Maryland: Chevy Chase.

Orrell, M., Yard, P., Handysides, J. & Schapira, R. (1999). Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community. *British Journal of Psychiatry, 174*, 409-412.

Overall, J. E. & Gorham, D. R. (1976). BPRS. Brief Psychiatric Rating Scale. In W. Guy (Hrsg.), *ECDEU assessment manual for psychopharmacology* (S. 157-169). Rockville: National Institute of Mental Health.

Page, A. C., Hooke, G. R. & Rutherford, E. M. (2001). Measuring mental health outcomes in a private psychiatric clinic: Health of the Nation Outcome Scales and Medical Outcomes Short Form SF-36. *Australian and New Zealand Journal of Psychiatry, 35*, 377-381.

Parker, G., O'Donnell, M., Hadzi-Pavlovic, D. & Proberts, M. (2002). Assessing outcome in community mental health patients: A comparative analysis of measures. *International Journal of Social Psychiatry, 48*(1), 11-19.

Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurment, 1*(3), 385-401.

Richter, D. (1999). Krankenhausbetriebsvergleich für psychiatrische Kliniken - Wie sinnvoll ist ein Vergleich der Behandlungsdauern nach Diagnose und Alter? *Gesundheitswesen, 61*, 227-233.

Richter, D. (2001). Die Dauer der stationären psychiatrischen Behandlung: Eine Übersicht über Methodik, Einflussfaktoren und Auswirkungen. *Fortschritte Neurologie Psychiatrie, 69*, 19-31.

Rochell, B. & Roeder, N. (2000). DRG-basierte Entgeltsysteme in Europa. In M. Arnold, M. Litsch & H. Schellschmidt (Hrsg.), *Krankenhaus-Report 2000* (S. 49-64). Stuttgart: Schattauer.

Rochell, B. & Roeder, N. (2002). DRGs als Grundlage der künftigen Krankenhausfinanzierung - Stand der Umsetzung und Einfluss auf die Rehabilitation. *Rehabilitation, 41*, 1-9.

Rock, D. & Preston, N. (2001). HoNOS: Is there any point in training clinicians? *Journal of Psychiatric and Mental Health Nursing, 8*, 405-409.

Roeder, N., Rochell, B. (2000). *Empirischer Vergleich von Patientenklassifikationssystemen auf der Grundlage von Diagnosis Related Groups (DRG) in der Herzchirurgie.* (Unveröffentlichter Zwischenbericht). Berlin: Deutsche Krankenhausgesellschaft; 2000.

Roick, C., Gärtner, A., Heider, D. & Angermeyer, M. C. (2002b). "Heavy user" psychiatrischer Versorgungsdienste: Ein Überblick über den Stand der Forschung. *Psychiatrische Praxis, 29*, 334-342.

Roick, C., Heider, D., Kilian, R. & Angermeyer, M. C. (2002a). Patienten mit starker Inanspruchnahme stationär-psychiatrischer Versorgung: Eine Analyse von Krankenkassendaten aus der Stadt Leipzig. *Psychiatrische Praxis, 29*, 343-349.

Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen. (2003). *Gutachten 2003: Finanzierung, Nutzerorientierung und Qualität*. Baden-Baden.

Sammet, I., Bohn, T., Leichsenring, F., Schauenburg, H. & Andreas, S. (in press). Self-evaluation of pathogenic beliefs: A pilot study basing on Control Mastery Theory. *Psychotherapy Research*.

Sartorius, N. & Kuyken, W. (1994). Translation of health status instruments. In J. Orley & W. Kuyken (Hrsg.), *Quality of life assessment: International perspectives* (S. 3-18). Heidelberg: Springer.

Saß, H., Wittchen, H.-U. & Zaudig, M. (1998). *Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Übersetzt nach der vierten Auflage des Diagnostic and Statistical Manual of Mental Disorders der American Psychiatric Association* (3. Aufl.). Göttingen: Hogrefe.

Schepank, H. (1995). *Der Beeinträchtigungs-Schwere-Score (BSS)*. Göttingen: Beltz Test.

Schepank, H. (2002). BSS: Der Beeinträchtigungs-Schwere-Score. In E. Brähler & B. Strauß (Hrsg.), *Diagnostische Verfahren in der Psychotherapie* (S. 68-72). Göttingen: Hogrefe-Verlag.

Scheytt, D., Kaiser, P. & Priebe, S. (1996). Behandlungsdauer und Fallkosten in unterschiedlichen stationären psychiatrischen Einrichtungen in Berlin. *Psychiatrische Praxis*, 10-14.

Schulz, H., Barghaan, D., Harfst, T. & Koch, U. (2004). *Gesundheitsberichterstattung des Bundes: Psychotherapeutische Versorgung*.

Schulz, H., Barghaan, D., Harfst, T. & Koch, U. (in Druck). *Gesundheitsberichtserstattung des Bundes: Psychotherapeutische Versorgung*. Berlin: Robert-Koch-Institut.

Schulz, H. & Koch, U. (2002). Zur stationären psychosomatisch-psychotherapeutischen Versorgung in Norddeutschland - Expertise zu Fragen des Bedarfs und zur Versorgungsstruktur. *Psychotherapie Psychosomatik Medizinische Psychologie, 52*, 244-247.

Schumacher, D. N., Namerow, J., Parker, B., Fox, P. & Kofie, V. (1986). Prospective payment for psychiatry - feasibility and impact. *New England Journal of Medicine, 315*, 1331-1336.

Schweizer, K. & Vach, W. (1999). Methode der Klassifikationsbäume (CART). In K. Schweizer (Hrsg.), *Methoden für die Analyse von Fragebogendaten* (S. 153-165). Göttingen: Hogrefe-Verlag.

Shepard, D. (1995). *The new direct marketing: How to implement a profit-driven database marketing strategy*. Homewood, Il.: Business One Irwin.

Spießl, H., Hübner-Liebermann, B., Binder, H. & Cording, C. (2002). "Heavy Users" in einer psychiatrischen Klinik - Eine Kohortenstudie mit 1811 Patienten über fünf Jahre. *Psychiatrische Praxis, 29*, 350-354.

Spitzer, C., Michels-Lucht, F., Siebel, U. & Freyberger, H. J. (2002). Die Strukturachse der operationalisierten psychodynamischen Diagnostik (OPD): Zusammenhänge mit soziodemographischen, klinischen und psychopathologischen Merkmalen sowie kategorialen Diagnosen. *Psychotherapie Psychosomatik Medizinische Psychologie, 52*, 392-397.

SPSS. (2001). *AnswerTree 3.0: Benutzerhandbuch*. Chicago: SPSS.

SPSS. (2004). *Answer Tree 3.1 for Windows*. Chicago: SPSS.

Stieglitz, R.-D. (1999). *Diagnostik und Klassifikation psychischer Störungen*. Göttingen: Hogrefe-Verlag.

Stieglitz, R.-D., Ahrens, B. & Freyberger, H. J. (2001). Fremdbeurteilungsverfahren. In R.-D. Stieglitz, U. Baumann & H. J. Freyberger (Hrsg.), *Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie* (Band 2. überarb. und erw. Auflage). Stuttgart: Thieme-Verlag.

Stieglitz, R.-D. & Freyberger, H. J. (2001). Selbstbeurteilungsverfahren. In R.-D. Stieglitz, U. Baumann & H. J. Freyberger (Hrsg.), *Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie* (Band 2. überarb. und erw. Auflage, S. 83-92). Stuttgart: Thieme-Verlag.

Stoskopf, C. & Horn, S. D. (1991). The Computerized Psychiatric Severity Index as a predictor of inpatient length of stay for psychoses. *Medical Care, 29*(3), 179-195.

Stoskopf, C. & Horn, S. D. (1992). Predicting length of stay for patients with psychoses. *Health Services Research, 26*(6), 743-766.

Taube, C., Lee, E. S. & Forthofer, R. N. (1984a). Diagnosis-Related Groups for mental disorders, alcoholism, and drug abuse: Evaluation and alternatives. *Hospital and Community Psychiatry, 35*(5), 452-455.

Taube, C., Lee, E. S. & Forthofer, R. N. (1984b). DRGs in Psychiatry. An empirical evaluation. *Medical Care, 22*, 597-610.

Tinsley, H. E. A. & Weiss, D. J. (1975). Interrater reliability and agreementof subjective judgements. *Journal of Counseling Psychology, 22*(4), 358-376.

Trauer, T. (2003). *Analysis of outcome measurement data from the four Victorian "Round One" agencies*.Unveröffentlichtes Manuskript, Melbourne, Australia.

Trauer, T., Callaly, T., Hantz, P., Little, J., Shields, R. & Smith, J. (1999). Health of the Nation Outcome Scales: Results of the Victorian field trial. *British Journal of Psychiatry, 174*, 380-388.

Trauer, T. (1999). The subscale structure of the Health of the Nation Outcome Scales (HoNOS). *Journal of Mental Health, 8*, 499-509.

Tritt, K., v. Heymann, F., Zaudig, M. & Loew, T. (2002). Ist die "diagnosebezogene Behandlungsdauer" eine homogene Einheit? *Psychotherapie Psychosomatik Medizinische Psychologie, 52*, 115.

Tuschen, K. H. (2001). BMG setzt Anreize für Einstieg 2003: Referentenentwurf eines DRG-Einführungsgesetzes. *f&w, 4*, 334-340.

Tuschen, K. H., Rau, F. & Braun, T. (2003). Das BMG sichert fristgerecht den Einstieg ins DRG-Fallpauschalensystem: Jetzt haben die Krankenhäuser die Wahl. *f&w*, 436-444.

Vassend, O. & Skrondal, A. (1999). The problem of structural indeterminancy in multidimensional symptom report instruments: The case of SCL-90-R. *Behaviour Research and Therapy, 37*, 685-701.

Verband Deutscher Rentenversicherungsträger. (2004). *VDR Statistik Rehabilitation des Jahres 2003*. Frankfurt a. M.: VDR.

von Heymann, F., Zaudig, M. & Tritt, K. (2003). Die diagnosebezogene Behandlungsdauer in der Psychosomatischen und Psychotherapeutischen Medizin: Eine homogene Größe? *Praxis Klinische Verhaltensmedizin und Rehabilitation, 62*, 209-221.

von Zerssen, D. (1976). *Die Befindlichkeits-Skala*. Weinheim: Beltz.

von Zerssen, D. (2002). Die Befindlichkeits-Skala (Bf-S). In E. Brähler, J. Schumacher & B. Strauß (Hrsg.), *Diagnostische Verfahren in der Psychotherapie* (S. 43-46). Göttingen: Hogrefe-Verlag.

Ware, J. E. (2000). SF- 36 Health survey update. *Spine, 25*(24), 3130- 3139.

Ware, J. E., Kosinski, M., Dewey, J. E. & Gandek, B. (2000). *How to score and interpret single-item health status measures: A manual for users of the SF-8 Health Survey*: Lincoln: Quality Metric.

WHO. (1992). *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organisation.

Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. B. G., Hadden, S. & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS): Research and development. *British Journal of Psychiatry, 172*, 11-18.

Wing, J. K., Curtis, R. H. & Beevor, A. S. (1996). *HoNOS: Health of the Nation Outcome Scales: Report on research and development July 1993-December 1995*. London: Royal College of Psychiatrists.

Wirtz, M. (2004). Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann. *Rehabilitation, 43*(2), 109-115.

Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen.* Göttingen: Hogrefe.

Young, W., Joyce, D. Z., Walrath, J. M., Schuchert, J. L., Hutton, M. A., Macioce, D. P., Patterson, M. S., Groetzinger, S., Bivens, G. D., Chiou, G. S., Emes, J. J., Hall, D. L., Winwood, J. P., Meanor, M. K., McLaughlin, K. M., Fonzi, P. J. & Woodburn, M. A. (1985). *Measuring the cost of care using patient management categories* (Band I-III). Pittsburgh: The Pittsburgh Research Institute, Center for Health Services Research.

Zielke, M., Dehmlow, A., Wülbeck, B. & Limbacher, K. (1997). Einflussfaktoren auf die Behandlungsdauer bei psychischen und psychosomatischen Erkrankungen in der stationären Verhaltenstherapie. *Praxis Klinische Verhaltensmedizin und Rehabilitation, 37*, 22-56.

Zobel, M., Missel, P., Bachmeier, R., Brünger, M., Funke, W., Herder, F., Kluger, H., Medenwaldt, J., Weissinger, V. & Wüst, G. (2004). Effektivität der stationären Suchtrehabilitation - FVS-Katamnese des Entlassjahrgangs 2001 von Fachkliniken für Alkohol- und Medikamentenabhängige. *Sucht aktuell, 11*, 11-20.

# 14 Tabellenverzeichnis

## 15 Abbildungsverzeichnis

# 16 Anhang

## 16.1 Materialien zur Einschätzung im HoNOS-D

### *16.1.1 Die deutsche Version der Health of the Nation Outcome Scales (HoNOS-D)*

## HoNOS-D Rating

Beurteilen Sie für jedes der 12 Items das jeweils **schwerste** Problem, das während der **letzten 7 Tage** aufgetreten ist.

**Legende**: Tragen Sie für jedes Item 0 - 4 oder 9 in die Kästchen ein.

| | | | |
|---|---|---|---|
| **0** | kein Problem | **1** | klinisch unbedeutendes Problem ohne Handlungsbedarf |
| **2** | leichtes Problem, aber eindeutig vorhanden | **3** | eher schweres Problem |
| **4** | schweres bis sehr schweres Problem | **9** | nicht bekannt / nicht anwendbar |

1) Überaktives, aggressives, Unruhe stiftendes oder agitiertes Verhalten ☐

2) Absichtliche Selbstverletzung ☐

3) Problematischer Alkoholkonsum oder Drogenkonsum ☐

4) Kognitive Probleme ☐

5) Probleme in Zusammenhang mit körperlicher Erkrankung oder Behinderung ☐

6) Probleme in Zusammenhang mit Halluzinationen und Wahnvorstellungen ☐

7) Gedrückte Stimmung ☐

8) Andere psychische und verhaltensbezogene Probleme ☐

*Falls andere psychische /verhaltensbezogene Probleme vorhanden, geben Sie die Art der Probleme an* ☐

| | |
|---|---|
| *A Phobisch* | *B Angst* |
| *C Zwangsgedanken/ -handlungen* | *D Psychische Belastung / Anspannung* |
| *E Dissoziativ* | *F Somatoform* |
| *G Essen* | *H Schlaf* |
| *I Sexuell* | *J Andere (Spezifizieren)* ________ |

9) Probleme mit Beziehungen ☐

10) Probleme mit alltäglichen Aktivitäten ☐

11) Probleme durch die Wohnbedingungen ☐

12) Probleme durch die Bedingungen im Beruf und im Alltag ☐

### *16.1.2 Das Glossar der HoNOS-D*




# Glossar für die Beurteilung der HoNOS-D

Arbeitsgruppe für Psychotherapie- und Versorgungsforschung
(AGPV)

Institut und Poliklinik für Medizinische Psychologie
Universitätsklinikum Hamburg-Eppendorf
Zentrum für Psychosoziale Medizin
Martinistr. 52, S 35
20246 Hamburg

Ansprechpartnerin:
Sylke Andreas
Diplom-Psychologin
Tel.: 040-42803-7705
Fax: 040-42803-4940
E-mail: sandreas@uke.uni-hamburg.de

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 1. Überaktives, aggressives, Unruhe stiftendes oder agitiertes Verhalten

** Beziehen Sie solche Verhaltensweisen aufgrund jeglicher Ursache ein ( z. B. Drogen, Alkohol, Demenz, Psychose, Depression usw.).*

** Beziehen Sie bizarres Verhalten, welches auf der Skala 6 eingeschätzt wird, **nicht** mit ein.*

0 Kein Problem dieser Art während des Bewertungszeitraums.

1 Reizbarkeit, Streitigkeiten, Ruhelosigkeit usw., ohne Handlungsbedarf.

2 Beinhaltet aggressive Gesten, Schubsen oder Belästigen von Anderen; Drohungen oder verbale Aggression; leichtere Sachbeschädigung (z. B. zerbrochene Tasse, zerbrochenes Fenster); ausgeprägte Hyperaktivität oder Agitiertheit.

3 körperlich aggressiv gegenüber anderen Personen oder Tieren (im Ausmaß Kodierung 4 unterschreitend); bedrohliches Auftreten; schwerere Überaktivität oder Zerstörung von Eigentum.

4 Mindestens ein schwerer körperlicher Angriff gegen andere Personen oder Tiere; Zerstörung von Eigentum (z. B. Brandstiftung); schwere Einschüchterung oder obszönes Verhalten.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 2. Absichtliche Selbstverletzung

* *Beziehen Sie versehentliche Selbstverletzung (z. B. aufgrund von Demenz oder schwerer Lernbehinderung)* ***nicht*** *mit ein; das kognitive Problem wird auf der Skala 4 und die Verletzung auf der Skala 5 eingeschätzt.*

* *Beziehen Sie Erkrankung oder Verletzung als direkte Folge von Drogen-/Alkoholkonsum* ***nicht*** *mit ein.*
  *(Das Ausmaß des Drogen- / Alkoholkonsums wird auf der Skala 3 eingeschätzt; mögliche körperliche Folgen des Drogen- / Alkoholkonsums wie z.B. Leberzirrhose oder Verletzungen, die durch Trunkenheit am Steuer verursacht werden, werden auf der Skala 5 eingeschätzt)*

0 Kein Problem dieser Art während des Bewertungszeitraums.

1 Flüchtige Gedanken, alles zu beenden, jedoch geringes Risiko während des Bewertungszeitraums; keine Selbstverletzung.

2 Mäßiges Risiko während des Bewertungszeitraums; schließt ungefährliche Selbstverletzung (z. B. Kratzen der Handgelenke) ein.

3 Mittleres bis schweres Risiko der absichtlichen Selbstverletzung während des Bewertungszeitraums; schließt vorbereitende Handlungen (z. B. das Sammeln von Tabletten) ein.

4 Schwerer Suizidversuch und/oder schwere absichtliche Selbstverletzung während des Bewertungszeitraums.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 3. Problematischer Alkoholkonsum oder Drogenkonsum

* *Beziehen Sie aggressives/destruktives Verhalten aufgrund von Alkohol- oder Drogenkonsum* ***nicht*** *mit ein. Dieses wurde auf der Skala 1 eingeschätzt.*

* *Beziehen Sie körperliche Erkrankung oder Behinderung aufgrund von Alkohol- oder Drogenkonsum* ***nicht*** *mit ein. Diese wird auf der Skala 5 eingeschätzt.*

0 Kein Problem dieser Art während des Bewertungszeitraums.

1 Gelegentlich übermäßiger Konsum, jedoch innerhalb der sozialen Norm.

2 Verlust der Kontrolle über das Trinken oder den Drogenkonsum, jedoch nicht schwer abhängig.

3 Ausgeprägtes Verlangen nach oder Abhängigkeit von Alkohol oder Drogen, mit häufigem Verlust der Kontrolle; Risikoverhalten unter dem Einfluss von Alkohol oder Drogen.

4 Durch das Alkohol-/Drogenproblem unfähig gemacht.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 4. Kognitive Probleme

* *Beziehen Sie Gedächtnis-, Orientierungs- und Verständnisprobleme in Zusammenhang mit irgendeiner Störung ein: Lernbehinderung, Demenz, Schizophrenie usw.*

* *Beziehen Sie vorübergehende Probleme (z. B. Kater) als Folge des Alkohol-/Drogenkonsums* ***nicht*** *mit ein. Diese werden auf der Skala 3 eingeschätzt.*

0 Kein Problem dieser Art während des Bewertungszeitraums.

1 Klinisch unbedeutende Probleme mit Gedächtnis oder Verständnis (z. B. vergisst gelegentlich Namen).

2 Leichte, aber eindeutig vorhandene Probleme (z. B. hat sich an einem bekannten Ort verlaufen oder erkennt eine bekannte Person nicht wieder); manchmal bei einfachen Entscheidungen verwirrt.

3 Ausgeprägte Desorientierung hinsichtlich Zeit, Ort oder Person; verwirrt durch Alltagsereignisse; Sprache ist manchmal unzusammenhängend; mentale Verlangsamung.

4 Schwere Desorientierung (z. B. erkennt Verwandte nicht wieder); unfallgefährdet; Sprache unverständlich; Bewusstseinstrübung oder Stupor.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 5. Probleme in Zusammenhang mit körperlicher Erkrankung oder Behinderung

* *Beziehen Sie Erkrankungen oder Behinderungen jedweder Ursache ein, welche die Bewegung einschränken oder verhindern, das Sehen oder Hören beeinträchtigen oder anderweitig die persönliche Funktionsfähigkeit beeinträchtigen.*

* *Beziehen Sie Nebenwirkungen von Medikamenten; Auswirkungen des Alkohol-/Drogenkonsums; körperliche Behinderung als Folge von Unfällen oder Selbstverletzung in Zusammenhang mit kognitiven Problemen, Trunkenheit am Steuer usw. ein.*

* *Beziehen Sie psychische oder verhaltensbezogene Probleme, welche auf der Skala 4 eingeschätzt werden, nicht mit ein.*

0 Keine körperlichen Gesundheitsprobleme während des Bewertungszeitraums.

1 Klinisch unbedeutende Gesundheitsprobleme während des Zeitraums (z. B. Erkältung, leichter Sturz usw.).

2 Körperliches Gesundheitsproblem schränkt die Mobilität und Aktivität leicht ein.

3 Mittlerer Grad der Aktivitätseinschränkung aufgrund eines körperlichen Gesundheitsproblems.

4 Schwere oder vollständige Funktionsunfähigkeit aufgrund von körperlichen Gesundheitsproblemen.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 6. Probleme in Zusammenhang mit Halluzinationen und Wahnvorstellungen

* *Beziehen Sie Halluzinationen und Wahnvorstellungen ungeachtet der Diagnose ein.*

* *Beziehen Sie merkwürdiges und bizarres Verhalten in Zusammenhang mit Halluzinationen oder Wahnvorstellungen ein.*

* *Beziehen Sie aggressive, destruktive oder überaktive Verhaltensweisen, die auf Halluzinationen oder Wahnvorstellungen zurückzuführen sind* ***nicht*** *mit ein. Diese werden auf der Skala 1 eingeschätzt.*

0 Kein Anzeichen von Halluzinationen oder Wahnvorstellungen während des Bewertungszeitraums.

1 Etwas merkwürdige oder exzentrische Überzeugungen, die nicht mit den kulturellen Normen übereinstimmen.

2 Wahnvorstellungen oder Halluzinationen (z. B Stimmen, Visionen) sind vorhanden, jedoch besteht geringes Leiden für den Patienten oder eine geringe Manifestation von bizarrem Verhalten, d.h. klinisch vorhanden, aber leicht.

3 Ausgeprägte Beschäftigung mit Wahnvorstellungen oder Halluzinationen, die starkes Leiden verursacht und/oder sich in offensichtlich bizarrem Verhalten manifestiert, d.h. eher schweres klinisches Problem.

4 Psychischer Zustand und Verhalten sind schwer und nachteilig durch Wahnvorstellungen oder Halluzinationen beeinträchtigt, mit schweren Auswirkungen auf den Patienten.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 7. Gedrückte Stimmung

* *Beziehen Sie Überaktivität oder Agitiertheit, welche auf der Skala 1 eingeschätzt werden, **nicht** mit ein.*

* *Beziehen Sie Suizidgedanken oder –versuche, welche auf der Skala 2 eingeschätzt werden, **nicht** mit ein.*

* *Beziehen Sie Wahnvorstellungen oder Halluzinationen, welche auf der Skala 6 eingeschätzt werden, **nicht** mit ein.*

0 Kein Problem in Zusammenhang mit gedrückter Stimmung während des Bewertungszeitraums.

1 Bedrückt; oder geringe Veränderungen in der Stimmung.

2 Leichte(s), jedoch eindeutig vorhandene(s) Depression und Leiden (z. B. Schuldgefühle; Verlust des Selbstwertgefühls).

3 Depression mit unangemessenen Selbstvorwürfen; zwanghaft beschäftigt mit Schuldgefühlen.

4 Schwere oder sehr schwere Depression mit Schuld oder Selbstanklage.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 8. Andere psychische und verhaltensbezogene Probleme

* *Schätzen Sie nur das schwerste klinische Problem ein, welches auf den Skalen 6 und 7 nicht berücksichtigt wurde.*

* *Spezifizieren Sie die Art des Problems, indem Sie den entsprechenden Buchstaben eintragen: A Phobisch; B Angst; C Zwangsgedanken/ -handlungen; D Psychische Belastung / Anspannung; E Dissoziativ; F Somatoform; G Essen; H Schlaf; I Sexuell; J Andere (Spezifizieren).*

0 Kein Anzeichen für irgendeines dieser Probleme während des Bewertungszeitraums.

1 Nur klinisch unbedeutende Probleme.

2 Ein Problem ist klinisch in leichter Ausprägung vorhanden (z. B. Patient hat ein gewisses Ausmaß an Kontrolle).

3 Gelegentlich schwerer Anfall oder Leiden, mit Verlust der Kontrolle (z. B. Patient muss sämtliche Angst hervorrufenden Situationen vermeiden, einen Nachbarn als Hilfe hinzuziehen usw.), d.h. eher schwere Ausprägung des Problems.

4 Schweres Problem beherrscht die meisten Aktivitäten.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 9. Probleme mit Beziehungen

** Schätzen Sie das schwerste Problem des Patienten in Zusammenhang mit aktivem oder passivem Rückzug aus sozialen Beziehungen und/oder nicht unterstützende, destruktive oder selbstschädigende Beziehungen ein.*

0 Kein bedeutendes Problem während des Bewertungszeitraums.

1 Geringe nicht-klinische Probleme.

2 Deutliches Problem beim Aufbau oder Aufrechterhalten von unterstützenden Beziehungen: Patient beklagt sich und/oder Probleme sind für andere offensichtlich.

3 Persistierendes größeres Problem aufgrund von aktivem oder passivem Rückzug aus sozialen Beziehungen und/oder Beziehungen, die geringen oder gar keinen Trost oder Unterstützung bieten.

4 Schwere und leidvolle soziale Isolation aufgrund der Unfähigkeit, sozial zu kommunizieren und/oder Rückzug aus sozialen Beziehungen.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 10. Probleme mit alltäglichen Aktivitäten

* *Schätzen Sie das allgemeine Funktionsniveau bei alltäglichen Aktivitäten ein (z. B. Probleme mit grundlegenden Aktivitäten der Selbstpflege, wie Essen, Waschen, Ankleiden, Morgentoilette; ebenso komplexe Fähigkeiten wie Haushaltsplanung, Organisieren einer Unterkunft, eines Berufs und der Freizeit, Mobilität und Benutzung von Transportmitteln, Einkaufen, Selbstentwicklung usw.).*

* *Beziehen Sie jeglichen Mangel an Motivation, Selbsthilfe-Möglichkeiten zu nutzen, mit ein, da dies zu einem insgesamt niedrigerem Funktionsniveau beiträgt.*

* *Beziehen sie den Mangel an Gelegenheiten, intakte Fähigkeiten und Fertigkeiten auszuüben nicht mit ein. Diese werden auf den Skalen 11-12 eingeschätzt.*

0 Kein Problem während des Bewertungszeitraums; gute Funktionsfähigkeit in allen Bereichen.

1 Nur klinisch unbedeutende Probleme (z. B. unordentlich, unorganisiert).

2 Angemessene Selbstpflege, jedoch erheblicher Leistungsmangel bei einer oder mehreren komplexen Fertigkeiten (siehe oben).

3 Erhebliches Problem in einem oder mehreren Bereichen der Selbstpflege (Essen, Waschen, Ankleidung, Morgentoilette) sowie starke Unfähigkeit, mehrere komplexe Fertigkeiten auszuüben.

4 Schwere Behinderung oder vollständige Unfähigkeit in allen oder nahezu allen Bereichen der Selbstpflege und komplexen Fertigkeiten.

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 11. Probleme durch die Wohnbedingungen

* *Schätzen Sie insgesamt die Schwere von Problemen mit der Qualität der Wohnbedingungen und der täglichen häuslichen Routine ein. Werden die grundlegenden Lebensnotwendigkeiten erfüllt (Heizung, Licht, Hygiene)? Wenn ja, gibt es Hilfe, die Behinderungen zu bewältigen und eine Auswahl an Möglichkeiten, Fähigkeiten zu verwenden und neue zu entwickeln?*

* *Schätzen Sie nicht den Grad der funktionellen Behinderung ein. Dieser wird auf der Skala 10 eingeschätzt.*

*** Beachte: Schätzen Sie die übliche Unterkunft des Patienten ein. Wenn in stationärer Behandlung, bewerten Sie die Unterkunft zu Hause. Wenn keine Informationen zur Verfügung stehen, tragen Sie 9 ein.**

0 Unterkunft und Wohnbedingungen sind annehmbar; hilfreich, um jegliche Behinderung, welche auf der Skala 10 eingeschätzt wird, auf dem geringst möglichen Niveau zu halten; und die Selbsthilfe unterstützend.

1 Unterkunft ist einigermaßen akzeptabel, obgleich geringfügige oder vorübergehende Probleme bestehen (z. B. kein optimaler Ort, nicht die bevorzugte Wahl, usw.)

2 Bedeutendes Problem mit einem oder mehreren Aspekten der Unterkunft und/oder der Verwaltung (z. B. beschränkte Entscheidungsfreiheit; Personal oder Personen des Haushalts wissen nicht, wie die Behinderung einzugrenzen ist oder wie der Einsatz oder die Entwicklung neuer oder intakter Fertigkeiten unterstützt werden kann).

3 Zahlreiche belastende Probleme mit der Unterkunft (z. B. einige elementare Notwendigkeiten sind nicht vorhanden); Die Ausstattung der Unterkunft bietet minimale oder gar keine Möglichkeiten, die Unabhängigkeit des Patienten zu verbessern.

4 Die Unterkunft ist inakzeptabel, wodurch die Probleme des Patienten verschlimmert werden (z. B. elementare Notwendigkeiten sind nicht vorhanden, dem Patienten droht Räumung oder „Obdachlosigkeit“ oder die Wohnbedingungen sind anderweitig nicht tragbar).

*Kodieren Sie mit 9 falls nicht bekannt / nicht anwendbar*

## 12. Probleme durch die Bedingungen im Beruf und im Alltag

* *Schätzen Sie das Ausmaß der Probleme mit der Qualität der Milieubedingungen insgesamt ein. Gibt es Hilfe, die Behinderungen zu bewältigen?*
*Und gibt es Möglichkeiten, die den Beruf und die Freizeit betreffenden Fähigkeiten und Aktivitäten aufrechtzuerhalten oder weiterzuentwickeln? Berücksichtigen Sie Faktoren, wie Stigmatisierung, Mangel an qualifiziertem Personal, Zugang zu unterstützenden Angeboten und Einrichtungen, z.B. Personalausstattung und Ausrüstung von Tageszentren, Werkstätten, sozialen Vereinen, usw.*

* *Schätzen Sie nicht das Niveau der funktionellen Behinderung selbst ein. Dieses wird auf der Skala 10 eingeschätzt.*

* **Beachte: Schätzen Sie die übliche Situation des Patienten ein. Wenn in stationärer Behandlung, schätzen Sie die Aktivitäten in der Zeit vor der Aufnahme ein.**
**Wenn keine Informationen zur Verfügung stehen, kodieren Sie mit 9.**

0 Milieu des Patienten ist akzeptabel: hilfreich, um jegliche Behinderung, welche auf der Skala 10 eingeschätzt wird, auf dem geringst möglichen Niveau zu halten; und die Selbsthilfe unterstützend.

1 geringfügige oder vorübergehende Probleme (z. B. verspäteter Erhalt von Überweisungen): angemessene Einrichtungen sind verfügbar, jedoch nicht immer zu den gewünschten Zeiten, usw.

2 Beschränkte Auswahl an Aktivitäten; Mangel an angemessener Toleranz (z. B. zu Unrecht verweigerter Zutritt zu öffentlichen Bibliotheken oder Bädern usw.); benachteiligt durch Fehlen einer permanenten Adresse; unzureichende Betreuung oder professionelle Unterstützung; hilfreiches Milieu verfügbar, jedoch nur für eine sehr begrenzte Stundenzahl.

3 Ausgeprägter Mangel an verfügbaren qualifizierten Dienstleistungen, die helfen, das Ausmaß der bestehenden Behinderung herabsetzen; keine Möglichkeiten, die intakten Fertigkeiten einzusetzen oder neue hinzuzufügen; unqualifizierte Betreuung, welche schwer zugänglich ist.

4 Mangel an irgendwelchen Gelegenheiten für Tagesaktivitäten verschlimmert das Problem des Patienten.

## 16.2 Fragebögen zur Messung der Ergebnisqualität

### *16.2.1 Fragebogen für die Therapeuten zu Beginn der Behandlung (Auswahl)*

Aufnahmedatum:______________________

## Bitte bei Aufnahme ausfüllen:

| ALLGEMEINE ANGABEN | | | | |
|---|---|---|---|---|
| **Art der Maßnahme** | ☐ ARB / AHB | ☐ stat. HV | ☐ sonstige | |
| **Rentner/in** | | ☐ ja | ☐ nein | |
| **Rentenantrag** | | ☐ ja | ☐ nein | |
| **Arbeitsfähigkeit bei Aufnahme** | | ☐ ja | ☐ nein | |
| **Kostenträger** | ☐ GKV<br>☐ private KV | ☐ LVA<br>☐ sonstige | ☐ BfA | ☐ Knappschaft (RV) |

| DIAGNOSEN | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Diagnosen nach ICD-10 (Text)** | **ICD-10 Schlüssel** | | | | | **Zusätze** | | | |
| | | | | | | **A** | **B** | **C** | **D** |
| 1. | | | | | | | | | |
| 2. | | | | | | | | | |
| 3. | | | | | | | | | |
| 4. | | | | | | | | | |
| 5. | | | | | | | | | |

| A: Diagnosezusatz | B: Sicherheit der Diagnose | C: Schweregrad | D: Behandlungsbedarf in Ihrer Reha-Klinik |
|---|---|---|---|
| 0 = kein Zusatz erforderlich<br>1 = z.Zt. erscheinungsfrei<br>2 = akuter Schub/Rezidiv<br>3 = chronisch progredient<br>4 = Zustand nach | 0 = gesichert<br>1 = fraglich, Verdacht auf ... | 0 = unauffällig<br>1 = leicht<br>2 = mittel<br>3 = schwer | 0 = nicht vorhanden<br>1 = gering<br>2 = schwer |

| **Chronifizierung** der Haupterkrankung (in Jahren) ca. | ☐< 1 J. ☐1-2 J. ☐3-5 J. ☐5-10 J. ☐> 10 Jahre<br>☐ nicht beurteilbar ☐ Akutereignis |
|---|---|

| **Reha-Motivation**<br>(1=sehr hohe Motivation; 6=keine Motivation) | sehr hohe Motivation | ☐1------☐2------☐3------☐4------☐5------☐6 | keine Motivation |
|---|---|---|---|

## Fragen zum allgemeinen Gesundheitszustand zu Behandlungsbeginn (SF-8)

| **1** | **Wie würden Sie den Gesundheitszustand des Patienten/der Patientin in der vergangenen Woche im Allgemeinen beschreiben?** | | | | |
|---|---|---|---|---|---|
| ausgezeichnet | sehr gut | gut | weniger gut | schlecht | sehr schlecht |
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 | ☐ 6 |

| **2** | **Wie sehr haben Probleme mit der körperlichen Gesundheit den Patienten/die Patientin in der vergangenen Woche bei normalen körperlichen Tätigkeiten eingeschränkt (zu Fuß gehen, Treppen steigen)?** | | | |
|---|---|---|---|---|
| überhaupt nicht | sehr wenig | mäßig | ziemlich | war zu körperlichen Tätigkeiten nicht in der Lage |
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |

| **3** | **Inwieweit hatte der Patient/die Patientin in der vergangenen Woche wegen seiner/ihrer körperlichen Gesundheit Schwierigkeiten bei der Ausübung seiner/ihrer täglichen Arbeit bzw. Tätigkeiten zu Hause und außer Haus?** | | | |
|---|---|---|---|---|
| überhaupt nicht | ein bisschen | mäßig | ziemlich | war zu alltäglichen Tätigkeiten nicht in der Lage |
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |

| **4** | **Wie stark waren die Schmerzen des Patienten/der Patientin in der vergangenen Woche?** | | | | |
|---|---|---|---|---|---|
| hatte keine Schmerzen | sehr leicht | leicht | mäßig | stark | sehr stark |
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 | ☐ 6 |

| **5** | **Wie viel Energie hatte der Patient/die Patientin in der vergangenen Woche?** | | | |
|---|---|---|---|---|
| sehr viel | ziemlich viel | mäßig viel | ein wenig | gar keine |
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |

| **6** | **Wie sehr haben die körperliche Gesundheit oder seelischen Probleme des Patienten/der Patientin in der vergangenen Woche seine/ihre normalen Kontakte zu Familienangehörigen oder Freunden eingeschränkt?** | | | |
|---|---|---|---|---|
| überhaupt nicht | sehr wenig | mäßig | ziemlich | war zu diesen Kontakten nicht in der Lage |
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |

| **7** | **Wie sehr haben dem Patienten/der Patientin seelische Probleme (z.B. Angst, Niedergeschlagenheit oder Reizbarkeit) in der vergangenen Woche zu schaffen gemacht?** | | | |
|---|---|---|---|---|
| überhaupt nicht | etwas | mäßig | ziemlich | sehr |
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |

| **8** | **Wie sehr haben die persönlichen oder seelischen Probleme des Patienten/der Patientin ihn/sie in der vergangen Woche daran gehindert, seine/ihre normalen Tätigkeiten im Beruf, in der Schule/im Studium oder andere alltägliche Tätigkeiten auszuüben?** | | | |
|---|---|---|---|---|
| überhaupt nicht | sehr wenig | etwas | ziemlich | konnte alltägliche Aktivitäten nicht ausführen |
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |

## HoNOS-D Rating

Beurteilen Sie für jedes der 12 Items das jeweils **schwerste** Problem, das während der **letzten 7 Tage** aufgetreten ist.

**Legende**: Tragen Sie für jedes Item 0 - 4 oder 9 in die Kästchen ein.

| | | | |
|---|---|---|---|
| 0 | kein Problem | 1 | unbedeutendes Problem ohne Handlungsbedarf |
| 2 | leichtes Problem, aber eindeutig vorhanden | 3 | eher schweres Problem |
| 4 | schweres bis sehr schweres Problem | 9 | unbekannt / nicht anwendbar |

1) Überaktives, aggressives, Unruhe stiftendes oder agitiertes Verhalten ☐

2) Absichtliche Selbstverletzung ☐

3) Problematischer Alkoholkonsum oder Drogenkonsum ☐

4) Kognitive Probleme ☐

5) Probleme in Zusammenhang mit körperlicher Erkrankung oder Behinderung ☐

6) Probleme in Zusammenhang mit Halluzinationen und Wahnvorstellungen ☐

7) Gedrückte Stimmung ☐

8) Andere psychische und verhaltensbezogene Probleme ☐

*Falls andere psychische /verhaltensbezogene Probleme vorhanden, geben Sie die Art der Probleme an* ☐

*A Phobisch* *B Angst*
*C Zwanghaft* *D Psychische Belastung / Anspannung*
*E Dissoziativ*
*F Somatoform* *G Essen*
*H Schlaf* *I Sexuell*
*J Andere (Spezifizieren)* ____________

9) Probleme mit Beziehungen ☐

10) Probleme mit alltäglichen Aktivitäten ☐

11) Probleme durch die Wohnbedingungen ☐

12) Probleme durch die Bedingungen im Beruf und im Alltag ☐

### *16.2.2 Fragebogen für die Therapeuten zum Ende der Behandlung (Auswahl)*

# Bitte bei Entlassung ausfüllen:

## Fragen zum allgemeinen Gesundheitszustand zu Behandlungsende (SF-8)

**1 Wie würden Sie den Gesundheitszustand des Patienten/der Patientin in der vergangenen Woche im Allgemeinen beschreiben?**

| ausgezeichnet | sehr gut | gut | weniger gut | schlecht | sehr schlecht |
|---|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 | ☐6 |

**2 Wie sehr haben Probleme mit der körperlichen Gesundheit den Patienten/die Patientin in der vergangenen Woche bei normalen körperlichen Tätigkeiten eingeschränkt (zu Fuß gehen, Treppen steigen)?**

| überhaupt nicht | sehr wenig | mäßig | ziemlich | war zu körperlichen Tätigkeiten nicht in der Lage |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**3 Inwieweit hatte der Patient/die Patientin in der vergangenen Woche wegen seiner/ihrer körperlichen Gesundheit Schwierigkeiten bei der Ausübung seiner/ihrer täglichen Arbeit bzw. Tätigkeiten zu Hause und außer Haus?**

| überhaupt nicht | ein bisschen | mäßig | ziemlich | war zu alltäglichen Tätigkeiten nicht in der Lage |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**4 Wie stark waren die Schmerzen des Patienten/der Patientin in der vergangenen Woche?**

| hatte keine Schmerzen | sehr leicht | leicht | mäßig | stark | sehr stark |
|---|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 | ☐6 |

**5 Wie viel Energie hatte der Patient/die Patientin in der vergangenen Woche?**

| sehr viel | ziemlich viel | mäßig viel | ein wenig | gar keine |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**6 Wie sehr haben die körperliche Gesundheit oder seelischen Probleme des Patienten/der Patientin in der vergangenen Woche seine/ihre normalen Kontakte zu Familienangehörigen oder Freunden eingeschränkt?**

| überhaupt nicht | sehr wenig | mäßig | ziemlich | war zu diesen Kontakten nicht in der Lage |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**7 Wie sehr haben dem Patienten/der Patientin seelische Probleme (z.B. Angst, Niedergeschlagenheit oder Reizbarkeit) in der vergangenen Woche zu schaffen gemacht?**

| überhaupt nicht | etwas | mäßig | ziemlich | sehr |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**8 Wie sehr haben die persönlichen oder seelischen Probleme des Patienten/der Patientin ihn/sie in der vergangen Woche daran gehindert, seine/ihre normalen Tätigkeiten im Beruf, in der Schule/im Studium oder andere alltägliche Tätigkeiten auszuüben?**

| überhaupt nicht | sehr wenig | etwas | ziemlich | konnte alltägliche Aktivitäten nicht ausführen |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

## HoNOS-D Rating

Beurteilen Sie für jedes der 12 Items das jeweils **schwerste** Problem, das während der **letzten 7 Tage** aufgetreten ist.

**Legende**: Tragen Sie für jedes Item 0 - 4 oder 9 in die Kästchen ein.

| | | | |
|---|---|---|---|
| 0 | kein Problem | 1 | unbedeutendes Problem ohne Handlungsbedarf |
| 2 | leichtes Problem, aber eindeutig vorhanden | 3 | eher schweres Problem |
| 4 | schweres bis sehr schweres Problem | 9 | unbekannt / nicht anwendbar |

1) Überaktives, aggressives, Unruhe stiftendes oder agitiertes Verhalten ☐

2) Absichtliche Selbstverletzung ☐

3) Problematischer Alkoholkonsum oder Drogenkonsum ☐

4) Kognitive Probleme ☐

5) Probleme in Zusammenhang mit körperlicher Erkrankung oder Behinderung ☐

6) Probleme in Zusammenhang mit Halluzinationen und Wahnvorstellungen ☐

7) Gedrückte Stimmung ☐

8) Andere psychische und verhaltensbezogene Probleme ☐

*Falls andere psychische /verhaltensbezogene Probleme vorhanden, geben Sie die Art der Probleme an* ☐

*A Phobisch* — *B Angst*
*C Zwanghaft* — *D Psychische Belastung / Anspannung*
*E Dissoziativ*
*F Somatoform* — *G Essen*
*H Schlaf* — *I Sexuell*
*J Andere (Spezifizieren)* ____________________

9) Probleme mit Beziehungen ☐

10) Probleme mit alltäglichen Aktivitäten ☐

11) Probleme durch die Wohnbedingungen ☐

12) Probleme durch die Bedingungen im Beruf und im Alltag ☐

| **ALLGEMEINE ANGABEN** | | | |
|---|---|---|---|
| **Dauer der Rehabilitation** | ________ **Tage** | | |
| **Arbeitsfähigkeit bei Entlassung** | **nur bei erwerbstätigen Pat.**<br>❑ ja ❑ nein | | |
| **Leistungsfähigkeit - bezogen auf die letzte Tätigkeit -** | | | |
| ❑ mind. 6 Std | ❑ 3 bis unter 6 Std. | ❑ weniger als 3 Std.s | ❑ keine Angabe erforderlich |
| **Leistungsfähigkeit - bezogen auf den allgemeinen Arbeitsmarkt -** (s.o.) | | | |
| ❑ mind. 6 Std | ❑ 3 bis unter 6 Std. | ❑ weniger als 3 Std. | ❑ keine Angabe erforderlich |
| **Motivation für weitere Reha-Maßnahmen (Nachsorge)**<br>(1=sehr hohe Motivation; 6=keine Motivation) | sehr hohe Motivation ☐1------☐2------☐3------☐4------☐5------☐6 keine Motivation | | |

Wie bewerten Sie das Ergebnis der Rehabilitationsbehandlung insgesamt?
(0 = sehr schlecht bis 10 = ausgezeichnet)

sehr schlecht [**0**] [**1**] [**2**] [**3**] [**4**] [**5**] [**6**] [**7**] [**8**] [**9**] [**10**] ausgezeichnet

**Vielen Dank für Ihre Angaben!**

### *16.2.3 Fragebogen für die Patienten zu Beginn der Therapie (Auswahl)*

## Angaben zur Person

| 1 | Ihr Geschlecht? | Ihr Alter? | Ihr Gewicht? | Ihre Größe? |
|---|---|---|---|---|
| | ☐1 weiblich<br>☐2 männlich | ________ Jahre | ________ kg | ________ cm |

**2 Welcher Nationalität gehören Sie an?**

☐1 Deutsch ☐2 Deutsche(r) Übersiedler(in) (Jahr der Übersiedlung: |__|__|__|__|) ☐3 andere Nationalität und zwar: ____________________

**3 Wie ist Ihr aktueller Familienstand?**

☐1 ledig
☐2 verheiratet
☐3 getrennt lebend
☐4 geschieden
☐5 verwitwet
☐6 wieder verheiratet

**4 In welcher Partnersituation leben Sie?**

☐1 kurzfristig kein(e) Partner(in) (bis zu 1 Jahr)
☐2 langfristig kein(e) Partner(in) (mehr als 1 Jahr)
☐3 wechselnde(r) Partner(innen)
☐4 feste(r) Partner(in) (Ehepartner(in))
☐5 feste(r) Partner(in) (nicht Ehepartner(in))

**5 Welchen höchsten Schulabschluss haben Sie?**

☐1 noch in der Schule
☐2 kein Schulabschluss
☐3 Sonderschulabschluss
☐4 Hauptschulabschluss
☐5 Realschulabschluss
☐6 Fachabitur/Abitur
☐9 sonstiges

**6 Welchen höchsten Berufsabschluss haben Sie?**

☐1 noch in der Berufsausbildung
☐2 Lehre/ Fachschule
☐3 Meister
☐4 Fachhochschule/Universität
☐5 ohne Berufsabschluss
☐9 sonstiges

**7 Bei welcher gesetzlichen Krankenkasse sind Sie versichert?**

- ☐1 Ortskrankenkasse
- ☐2 Innungskrankenkasse
- ☐3 Betriebskrankenkasse
- ☐4 Ersatzkrankenkasse

**8 Welche berufliche Position nehmen Sie gegenwärtig ein?**
**(falls Sie arbeitslos bzw. berentet sind, geben Sie bitte Ihre letzte Position an)**

| Arbeiter(in): | Angestellte(r) | 1. Beamtin/Beamter |
|---|---|---|
| ☐1 ungelernte(r), angelernte(r) | ☐4 einfacher Ang. | ☐8 einfacher Dienst (A1-4) |
| | ☐5 mittlerer Ang. | ☐9 mittlerer Dienst (A5-8) |
| ☐2 Facharbeiter(in) | ☐6 höherer Ang. | ☐10 gehobener Dienst (A9-13) |
| ☐3 Meister(in) | ☐7 hochqualifizierter / ltd. Ang. | ☐11 höherer Dienst (A14-16) |

**Selbständige(r)**

- ☐12 kleiner Handwerks-, Landwirts-, Gewerbebetrieb
- ☐13 mittlerer Handwerks-, Landwirts-, Gewerbebetrieb
- ☐14 Akademiker(in), Freiberufler(in), Unternehmer(in) (größerer Betrieb)
- ☐15 **sonstiges**
- ☐16 **nie erwerbstätig**

**9 Wie ist Ihre aktuelle berufliche Situation?**

- ☐1 berufstätig, Vollzeit
- ☐2 berufstätig, Teilzeit
- ☐3 berufstätig, gelegentlich
- ☐4 mithelfender Familienangehöriger, nicht berufstätig
- ☐5 Hausfrau/-mann, nicht berufstätig
- ☐6 Ausbildung
- ☐7 Wehr-/Zivildienst/freiw. soz. Jahr
- ☐8 beschützt beschäftigt
- ☐9 arbeitslos gemeldet
- ☐10 Erwerbs-/Berufsunfähigkeitsrente, Erwerbsminderungsrente
  seit: .|__|__|.|__|__| (Monat Jahr)
- ☐11 Frührente, Frühpension
  seit: .|__|__|.|__|__| (Monat Jahr)
- ☐12 Altersrente/Pension
- ☐13 Witwen-/Witwer-Rente
- ☐14 anderweitig ohne berufliche Beschäftigung
- ☐15 unbekannt/unklar

**10 Sind Sie derzeitig arbeitsunfähig/krankgeschrieben?**

- ☐1 ja → seit wie vielen Wochen ununterbrochen? |__|__|__| (999 = unbekannt)
- ☐0 nein

| 11 | Haben Sie einen Antrag auf Erwerbsminderungsrente gestellt? | |
|---|---|---|
| ☐1 ja | → Wann haben Sie den Antrag gestellt? .|__|__|.|__|__| (Monat Jahr) ................................................ | |
| ☐0 nein | → Beabsichtigen Sie in naher Zukunft, einen Antrag auf Erwerbsminderungsrente zu stellen? | ☐1 ja ☐0 nein |

## Psychische und körperliche Beschwerden vor der stationären psychotherapeutische Behandlung

**21 Wie häufig hatten Sie in den letzten vier Wochen Selbstmordgedanken?**

| gar nicht | selten | mittel | oft | sehr oft |
|---|---|---|---|---|
| ☐0 | ☐1 | ☐2 | ☐3 | ☐4 |

**22 Wie häufig haben Sie in den letzten vier Wochen selbstverletzendes Verhalten wie schneiden, brennen, würgen etc. durchgeführt?**

| gar nicht | bis zu 1x pro Woche | an 2-6 Tagen pro Woche | täglich | mehrmals täglich |
|---|---|---|---|---|
| ☐0 | ☐1 | ☐2 | ☐3 | ☐4 |

**31 An wie vielen Tagen nahmen Sie in den letzten vier Wochen folgende Medikamente ein?**

| | **nie** | an **1-3 Tagen pro Monat**, Dosierung nach Bedarf | an **1-2 Tagen pro Woche**, Dosierung nach Bedarf | an **3-7 Tagen pro Woche**, Dosierung nach Bedarf | **täglich, in fester Dosierung** |
|---|---|---|---|---|---|
| **Schmerzmittel** | ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |
| **Beruhigungsmittel** | ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |
| **Schlafmittel** | ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |
| **Antidepressiva** | ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |
| **Neuroleptika** | ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |
| **Aufputschmittel** | ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

| | | | |
|---|---|---|---|
| **32** | **Rauchen Sie?** | ☐1 ja | ☐0 nein |
| | *Falls ja, wie viel?* **Zigaretten:** | ________ | Stück pro Tag |
| | **sonstiges (Pfeife, Zigarre, Zigarillo):** | ________ | Stück pro Tag |

**35** Sie finden nachstehend eine Liste von **Problemen und Beschwerden**, die man manchmal hat. Bitte lesen Sie sich jede Frage sorgfältig durch und entscheiden Sie, wie sehr Sie **in den letzten sieben Tagen** durch diese Beschwerden gestört oder bedrängt worden sind. Machen Sie bitte hinter jeder Frage nur ein Kreuz in das Kästchen mit der für Sie am besten zutreffenden Antwort. Streichen Sie versehentliche Antworten deutlich durch und kreuzen Sie danach das richtige Kästchen an.
**Bitte beantworten Sie jede Frage!**

**Wie sehr litten Sie in den letzten sieben Tagen unter ...?**

| | | überhaupt nicht | ein wenig | ziemlich | stark | sehr stark |
|---|---|---|---|---|---|---|
| 1 | Furcht auf offenen Plätzen oder auf der Straße | 0 | 1 | 2 | 3 | 4 |
| 2 | Befürchtungen, wenn Sie alleine aus dem Haus gehen | 0 | 1 | 2 | 3 | 4 |
| 3 | Selbstvorwürfen über bestimmte Dinge | 0 | 1 | 2 | 3 | 4 |
| 4 | dem Gefühl, dass es Ihnen schwer fällt, etwas anzufangen | 0 | 1 | 2 | 3 | 4 |
| 5 | Schwermut | 0 | 1 | 2 | 3 | 4 |
| 6 | Muskelschmerzen (Muskelkater, Gliederreißen) | 0 | 1 | 2 | 3 | 4 |
| 7 | Furcht vor Fahrten in Bus, Straßenbahn, U-Bahn oder Zug | 0 | 1 | 2 | 3 | 4 |
| 8 | Taubheit oder Kribbeln in einzelnen Körperteilen | 0 | 1 | 2 | 3 | 4 |
| 9 | einem Gefühl der Hoffnungslosigkeit angesichts der Zukunft | 0 | 1 | 2 | 3 | 4 |
| 10 | Schwächegefühl in einzelnen Körperteilen | 0 | 1 | 2 | 3 | 4 |
| 11 | Schweregefühl in den Armen oder den Beinen | 0 | 1 | 2 | 3 | 4 |
| 12 | Einsamkeitsgefühlen, selbst wenn Sie in Gesellschaft sind | 0 | 1 | 2 | 3 | 4 |
| 13 | dem Gefühl, wertlos zu sein | 0 | 1 | 2 | 3 | 4 |
| 14 | der Furcht, in der Öffentlichkeit in Ohnmacht zu fallen | 0 | 1 | 2 | 3 | 4 |

**36** **Fragen zu Ihrem Befinden**

Bitte kreuzen Sie bei den folgenden Aussagen die Antwort an, die Ihrem Befinden während der letzten Woche am besten entspricht/entsprochen hat.

| Antworten: | 0 | selten oder überhaupt nicht | (weniger als 1 Tag) |
|---|---|---|---|
| | 1 | manchmal | (1 bis 2 Tage lang) |
| | 2 | öfters | (3 bis 4 Tage lang) |
| | 3 | meistens, die ganze Zeit | (5 bis 7 Tage lang) |

**Während der letzten Woche ...**

| | selten 0 | manchmal 1 | öfters 2 | meistens 3 |
|---|---|---|---|---|
| 1. haben mich Dinge beunruhigt, die mir sonst nichts ausmachen | ☐0 | ☐1 | ☐2 | ☐3 |
| 2. konnte ich meine trübsinnige Laune nicht loswerden, obwohl mich meine Freunde/ Familie versuchten, aufzumuntern | ☐0 | ☐1 | ☐2 | ☐3 |
| 3. hatte ich Mühe, mich zu konzentrieren | ☐0 | ☐1 | ☐2 | ☐3 |
| 4. war ich deprimiert/ niedergeschlagen | ☐0 | ☐1 | ☐2 | ☐3 |
| 5. war alles anstrengend für mich | ☐0 | ☐1 | ☐2 | ☐3 |
| 6. dachte ich, mein Leben ist ein einziger Fehlschlag | ☐0 | ☐1 | ☐2 | ☐3 |
| 7. hatte ich Angst | ☐0 | ☐1 | ☐2 | ☐3 |
| 8. habe ich schlecht geschlafen | ☐0 | ☐1 | ☐2 | ☐3 |
| 9. war ich fröhlich gestimmt | ☐0 | ☐1 | ☐2 | ☐3 |
| 10. habe ich weniger als sonst geredet | ☐0 | ☐1 | ☐2 | ☐3 |
| 11. fühlte ich mich einsam | ☐0 | ☐1 | ☐2 | ☐3 |
| 12. habe ich das Leben genossen | ☐0 | ☐1 | ☐2 | ☐3 |
| 13. war ich traurig | ☐0 | ☐1 | ☐2 | ☐3 |
| 14. hatte ich das Gefühl, dass mich die Leute nicht leiden können | ☐0 | ☐1 | ☐2 | ☐3 |
| 15. konnte ich mich zu nichts aufraffen | ☐0 | ☐1 | ☐2 | ☐3 |

| 37 | Fragen zum Allgemeinen Gesundheitszustand |
|---|---|

**1 Wie würden Sie Ihren Gesundheitszustand in der vergangenen Woche im Allgemeinen beschreiben?**

| ausgezeichnet | sehr gut | gut | weniger gut | schlecht | sehr schlecht |
|---|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 | ☐6 |

**2 Wie sehr haben Probleme mit der körperlichen Gesundheit Sie in der vergangenen Woche bei normalen körperlichen Tätigkeiten eingeschränkt (zu Fuß gehen, Treppen steigen)?**

| überhaupt nicht | sehr wenig | mäßig | ziemlich | war zu körperlichen Aktivitäten nicht in der Lage |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**3 Inwieweit hatten Sie in der vergangenen Woche wegen Ihrer körperlichen Gesundheit Schwierigkeiten bei der Ausübung Ihrer täglichen Arbeit zu Hause und außer Haus?**

| überhaupt nicht | ein bisschen | mäßig | ziemlich | war zu alltäglicher Arbeit nicht in der Lage |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**4 Wie stark waren Ihre Schmerzen in der vergangenen Woche?**

| Ich hatte keine Schmerzen | sehr leicht | leicht | mäßig | stark | sehr stark |
|---|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 | ☐6 |

**5 Wie viel Energie hatten Sie in der vergangenen Woche?**

| sehr viel | ziemlich viel | mäßig viel | ein wenig | gar keine |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**6 Wie sehr haben Ihre körperliche Gesundheit oder seelischen Probleme in der vergangenen Woche Ihre normalen Kontakte zu Familienangehörigen oder Freunden eingeschränkt?**

| überhaupt nicht | sehr wenig | mäßig | ziemlich | war zu diesen Kontakten nicht in der Lage |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**7 Wie sehr haben Ihnen in der vergangenen Woche seelische Probleme (z.B. Angst, Niedergeschlagenheit oder Reizbarkeit) zu schaffen gemacht?**

| überhaupt nicht | etwas | mäßig | ziemlich | sehr |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

**8 Wie sehr haben Ihre persönlichen oder seelischen Probleme Sie in der vergangenen Woche daran gehindert, Ihre normalen Tätigkeiten im Beruf, in der Schule/im Studium oder andere alltägliche Tätigkeiten auszuüben?**

| überhaupt nicht | sehr wenig | mäßig | ziemlich | war dazu nicht in der Lage |
|---|---|---|---|---|
| ☐1 | ☐2 | ☐3 | ☐4 | ☐5 |

39

# Fragen zu Beziehungen zu anderen

Nachfolgend finden Sie eine Liste von Schwierigkeiten, die im Zusammenhang mit Beziehungen zu anderen berichtet werden. Bitte lesen Sie diese Liste durch und überlegen Sie, ob die einzelnen Schwierigkeiten für Sie ein Problem darstellten und zwar bezogen auf irgendeine Person, die in Ihrem Leben eine bedeutsame Rolle spielte.
Kreuzen Sie bitte für jedes Problem die Ziffer an, die beschreibt, wie sehr sie darunter gelitten haben.

**Teil I. Die nachstehenden Aspekte können im Umgang mit anderen schwierig sein.**

| **Es fällt mir schwer ...** | nicht | wenig | mittelmäßig | ziemlich | sehr |
|---|---|---|---|---|---|
| 1. anderen Menschen zu vertrauen | 0 | 1 | 2 | 3 | 4 |
| 2. anderen gegenüber „Nein“ zu sagen | 0 | 1 | 2 | 3 | 4 |
| 3. mich Gruppen anzuschließen | 0 | 1 | 2 | 3 | 4 |
| 4. bestimmte Dinge für mich zu behalten | 0 | 1 | 2 | 3 | 4 |
| 5. andere wissen zu lassen, was ich will | 0 | 1 | 2 | 3 | 4 |
| 6. jemanden zu sagen, dass er mich nicht weiter belästigen soll | 0 | 1 | 2 | 3 | 4 |
| 7. mich fremden Menschen vorzustellen | 0 | 1 | 2 | 3 | 4 |
| 8. andere mit anstehenden Problemen zu konfrontieren | 0 | 1 | 2 | 3 | 4 |
| 9. mich gegenüber jemand anderem zu behaupten | 0 | 1 | 2 | 3 | 4 |
| 10. andere wissen zu lassen, dass ich wütend bin | 0 | 1 | 2 | 3 | 4 |
| 11. eine langfristige Verpflichtung gegenüber anderen einzugehen | 0 | 1 | 2 | 3 | 4 |
| 12. jemandem gegenüber die „Chef-Rolle“ einzunehmen | 0 | 1 | 2 | 3 | 4 |
| 13. anderen gegenüber aggressiv zu sein, wenn es die Lage erfordert | 0 | 1 | 2 | 3 | 4 |
| 14. mit anderen etwas zu unternehmen | 0 | 1 | 2 | 3 | 4 |
| 15. anderen Menschen meine Zuneigung zu zeigen | 0 | 1 | 2 | 3 | 4 |
| 16. mit anderen zurechtzukommen | 0 | 1 | 2 | 3 | 4 |
| 17. die Ansichten eines anderen zu verstehen | 0 | 1 | 2 | 3 | 4 |
| 18. meine Gefühle anderen gegenüber frei heraus zu äußern | 0 | 1 | 2 | 3 | 4 |
| 19. wenn nötig, standfest zu sein | 0 | 1 | 2 | 3 | 4 |

| Es fällt mir schwer ... | | nicht | wenig | mittelmäßig | ziemlich | sehr |
|---|---|---|---|---|---|---|
| 20. | ein Gefühl von Liebe für jemanden zu empfinden | 0 | 1 | 2 | 3 | 4 |
| 21. | anderen Grenzen zu setzen | 0 | 1 | 2 | 3 | 4 |
| 22. | jemand anderen in seinen Lebenszielen zu unterstützen | 0 | 1 | 2 | 3 | 4 |
| 23. | mich anderen nahe zu fühlen | 0 | 1 | 2 | 3 | 4 |
| 24. | mich wirklich um die Probleme anderer zu kümmern | 0 | 1 | 2 | 3 | 4 |
| 25. | mich mit jemand anderem zu streiten | 0 | 1 | 2 | 3 | 4 |
| 26. | alleine zu sein | 0 | 1 | 2 | 3 | 4 |
| 27. | jemandem ein Geschenk zu machen | 0 | 1 | 2 | 3 | 4 |
| 28. | mir auch gegenüber den Menschen Ärger zu gestatten, die ich mag | 0 | 1 | 2 | 3 | 4 |
| 29. | die Bedürfnisse eines anderen über meine eigenen zu stellen | 0 | 1 | 2 | 3 | 4 |
| 30. | mich aus den Angelegenheiten anderer herauszuhalten | 0 | 1 | 2 | 3 | 4 |
| 31. | Anweisungen von Personen entgegenzunehmen, die mir vorgesetzt sind | 0 | 1 | 2 | 3 | 4 |
| 32. | mich über das Glück eines anderen Menschen zu freuen | 0 | 1 | 2 | 3 | 4 |
| 33. | andere zu bitten, mit mir etwas zu unternehmen | 0 | 1 | 2 | 3 | 4 |
| 34. | mich über andere zu ärgern | 0 | 1 | 2 | 3 | 4 |
| 35. | mich zu öffnen und meine Gefühle jemand anderem mitzuteilen | 0 | 1 | 2 | 3 | 4 |
| 36. | jemand anderem zu verzeihen, nachdem ich ärgerlich war | 0 | 1 | 2 | 3 | 4 |
| 37. | mein eigenes Wohlergehen nicht aus dem Auge zu verlieren, wenn jemand anderes in Not ist | 0 | 1 | 2 | 3 | 4 |
| 38. | fest und bestimmt zu bleiben, ohne mich darum zu kümmern, ob ich die Gefühle anderer verletze | 0 | 1 | 2 | 3 | 4 |
| 39. | selbstbewusst zu sein, wenn ich mit anderen zusammen bin | 0 | 1 | 2 | 3 | 4 |

| Teil II. Die nachstehenden Aspekte kann man im Übermaß tun. | nicht | wenig | mittelmäßig | ziemlich | sehr |
|---|---|---|---|---|---|
| 40. Ich streite mich zuviel mit anderen | 0 | 1 | 2 | 3 | 4 |
| 41. Ich fühle mich zu sehr für die Lösung der Probleme anderer verantwortlich | 0 | 1 | 2 | 3 | 4 |
| 42. Ich lasse mich zu leicht von anderen überreden | 0 | 1 | 2 | 3 | 4 |
| 43. Ich öffne mich anderen zu sehr | 0 | 1 | 2 | 3 | 4 |
| 44. Ich bin zu unabhängig | 0 | 1 | 2 | 3 | 4 |
| 45. Ich bin gegenüber anderen zu aggressiv | 0 | 1 | 2 | 3 | 4 |
| 46. Ich bemühe mich zu sehr, anderen zu gefallen | 0 | 1 | 2 | 3 | 4 |
| 47. Ich spiele zu oft den Clown | 0 | 1 | 2 | 3 | 4 |
| 48. Ich lege zuviel Wert darauf, beachtet zu werden | 0 | 1 | 2 | 3 | 4 |
| 49. Ich vertraue anderen nicht | 0 | 1 | 2 | 3 | 4 |
| 50. Ich bin zu sehr darauf aus, andere zu kontrollieren | 0 | 1 | 2 | 3 | 4 |
| 51. Ich stelle zu oft die Bedürfnisse anderer über meine eigenen | 0 | 1 | 2 | 3 | 4 |
| 52. Ich versuche zu sehr, andere zu verändern | 0 | 1 | 2 | 3 | 4 |
| 53. Ich bin zu leichtgläubig | 0 | 1 | 2 | 3 | 4 |
| 54. Ich bin anderen gegenüber zu großzügig | 0 | 1 | 2 | 3 | 4 |
| 55. Ich habe vor anderen zuviel Angst | 0 | 1 | 2 | 3 | 4 |
| 56. Ich bin anderen gegenüber zu misstrauisch | 0 | 1 | 2 | 3 | 4 |
| 57. Ich beeinflusse andere zu sehr, um zu bekommen, was ich will | 0 | 1 | 2 | 3 | 4 |
| 58. Ich erzähle anderen zu oft persönliche Dinge | 0 | 1 | 2 | 3 | 4 |
| 59. Ich streite zu oft mit anderen | 0 | 1 | 2 | 3 | 4 |
| 60. Ich halte mir andere zu sehr auf Distanz | 0 | 1 | 2 | 3 | 4 |
| 61. Ich lasse mich von anderen zu sehr ausnutzen | 0 | 1 | 2 | 3 | 4 |
| 62. Ich bin vor anderen Menschen zu verlegen | 0 | 1 | 2 | 3 | 4 |
| 63. Die Not eines anderen Menschen berührt mich zu sehr | 0 | 1 | 2 | 3 | 4 |
| 64. Ich möchte mich zu sehr an anderen rächen | 0 | 1 | 2 | 3 | 4 |

## • Medizinische und psychosoziale Maßnahmen vor der stationären psychotherapeutischen Behandlung

| 40 | • **Haben Sie in den letzten 12 Monaten vor der jetzigen stationären Behandlung an einer oder mehreren der folgenden psychosozialen Behandlungsmaßnahmen teilgenommen?** | | | | |
|---|---|---|---|---|---|
| | • | nein | 1 bis 5 Termine | 6 bis 25 Termine | mehr als 25 Termine |
| | ambulante Psychotherapie (Einzel) | ☐0 | ☐1 | ☐2 | ☐3 |
| | ambulante Gruppentherapie | ☐0 | ☐1 | ☐2 | ☐3 |
| | Familientherapie/Paartherapie | ☐0 | ☐1 | ☐2 | ☐3 |
| | ambulante psychiatrische Behandlung | ☐0 | ☐1 | ☐2 | ☐3 |
| | Tagesklinik | ☐0 | ☐1 | ☐2 | ☐3 |
| | Beratungsstelle | ☐0 | ☐1 | ☐2 | ☐3 |
| | Selbsthilfegruppe | ☐0 | ☐1 | ☐2 | ☐3 |
| | sonstiges, und zwar: ________________________ | | ☐1 | ☐2 | ☐3 |

| 41 | **Sind in den letzten 12 Monaten stationär, das heißt in einem Krankenhaus/ einer Klinik, behandelt worden?** | | | |
|---|---|---|---|---|
| | | nein | ja | Wie viele Krankenhaustage insgesamt (in Tagen)? |
| | psychotherapeutische oder psychiatrische Krankenhausaufenthalte | ☐0 | ☐1 ⇒ | _________ (Tage) |
| | sonstige Krankenhausaufenthalte | ☐0 | ☐1 ⇒ | _________ (Tage) |

## • Erwerbstätigkeit

**42a** • **Für Berufstätige/ arbeitslos Gemeldete:**

**Zeiten von Arbeitsunfähigkeit in den letzten 6 Monaten:**

Ich war in den letzten 6 Monaten insgesamt ca. _________ Wochen arbeitsunfähig.

Bitte zählen Sie alle Tage/ Wochen zusammen, an denen Sie wegen Krankheit nicht zur Arbeit konnten, auch wenn Sie nicht ärztlich krankgeschrieben waren. Falls Sie die Zahl nicht wissen, schätzen Sie möglichst genau!

**42b** • **Nur für Nicht-Berufstätige (Rentner/-in, Student/-in, Hausfrau/-mann, etc.):**

Ich war in den letzten 6 Monaten insgesamt ca. _________ Wochen so krank, dass ich meine Aufgaben nicht wie gewohnt erledigen konnte.

Falls Sie die Zahl nicht wissen, schätzen Sie möglichst genau!

***Wir danken Ihnen sehr herzlich für Ihre Mitarbeit!***

***Sehen Sie jetzt bitte den Fragebogen nochmals durch und prüfen Sie, ob Sie auch jede Frage beantwortet haben.***

## 16.3 Analysen zu den Verteilungseigenschaften der eingesetzten standardisierten Items

### *16.3.1 Verteilungseigenschaften der selbst eingeschätzten Items auf nicht-metrischem Skalenniveau*

Tabelle 40: Häufigkeitsverteilung des Prädiktors „Partnersituation“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| kurzfristig kein Partner | 224 | 11,7 | 11,7 | 11,7 |
| langfristig kein Partner | 456 | 23,8 | 23,8 | 35,5 |
| wechselnde Partner | 34 | 1,8 | 1,8 | 37,2 |
| fester Ehepartner | 664 | 34,6 | 34,6 | 71,8 |
| fester Partner (nicht Ehe-) | 261 | 13,6 | 13,6 | 85,5 |
| fehlend | 279 | 14,5 | 14,5 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 41: Häufigkeitsverteilung des Prädiktors „höchster beruflicher Abschluss“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| noch in der Berufsausbildung | 77 | 4,0 | 4,0 | 4,0 |
| Lehre/ Fachschule | 806 | 42,0 | 42,0 | 46,0 |
| Meister | 58 | 3,0 | 3,0 | 49,1 |
| Fachhochschule/Universität | 297 | 15,5 | 15,5 | 64,5 |
| ohne Berufsabschluss | 222 | 11,6 | 11,6 | 76,1 |
| sonstiges | 172 | 9,0 | 9,0 | 85,1 |
| fehlend | 286 | 14,9 | 14,9 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 42: Häufigkeitsverteilung des Prädiktors „höchster Schulabschluss“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| noch in der Schule | 72 | 3,8 | 3,8 | 3,8 |
| kein Schulabschluss | 30 | 1,6 | 1,6 | 5,3 |
| Sonderschulabschluss | 17 | ,9 | ,9 | 6,2 |
| Hauptschulabschluss | 551 | 28,7 | 28,7 | 34,9 |
| Realschulabschluss | 569 | 29,7 | 29,7 | 64,6 |
| Fachabitur/ Abitur | 513 | 26,7 | 26,7 | 91,3 |
| sonstiges | 66 | 3,4 | 3,4 | 94,8 |
| fehlend | 100 | 5,2 | 5,2 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 43: Häufigkeitsverteilung des Prädiktors „aktuelle berufliche Situation“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| berufstätig, Vollzeit | 543 | 28,3 | 28,3 | 28,3 |
| berufstätig, Teilzeit | 274 | 14,3 | 14,3 | 42,6 |
| berufstätig, gelegentlich | 39 | 2,0 | 2,0 | 44,6 |
| mithelfender Familienangehöriger, nicht berufstätig | 6 | ,3 | ,3 | 44,9 |
| Hausfrau/-mann, nicht berufstätig | 152 | 7,9 | 7,9 | 52,9 |
| Ausbildung | 102 | 5,3 | 5,3 | 58,2 |
| Wehr-/Zivildienst/freiw. soz. Jahr | 1 | ,1 | ,1 | 58,2 |
| beschützt beschäftigt | 4 | ,2 | ,2 | 58,4 |
| arbeitslos gemeldet | 232 | 12,1 | 12,1 | 70,5 |
| Erwerbs-/Berufsunfähigkeitsrente, Erwerbsminderungsrente | 109 | 5,7 | 5,7 | 76,2 |
| Frührente, Frühpension | 15 | ,8 | ,8 | 77,0 |
| Altersrente/Pension | 98 | 5,1 | 5,1 | 82,1 |
| Witwen-/Witwer-Rente | 22 | 1,1 | 1,1 | 83,3 |
| 13 | 1 | ,1 | ,1 | 83,3 |
| anderweitig ohne berufliche Beschäftigung | 84 | 4,4 | 4,4 | 87,7 |
| unbekannt/unklar | 52 | 2,7 | 2,7 | 90,4 |
| fehlend | 184 | 9,6 | 9,6 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 44: Häufigkeitsverteilung des Prädiktors „gegenwärtige berufliche Position“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| ungelernt, angelernt | 188 | 9,8 | 9,8 | 9,8 |
| Facharbeiter(in) | 111 | 5,8 | 5,8 | 15,6 |
| Meister(in) | 9 | ,5 | ,5 | 16,1 |
| einfacher Angestellter | 342 | 17,8 | 17,8 | 33,9 |
| mittlerer Angestellter | 327 | 17,0 | 17,0 | 51,0 |
| höherer Angestellter | 124 | 6,5 | 6,5 | 57,5 |
| hochqualifizierter / ltd. Angestellter. | 64 | 3,3 | 3,3 | 60,8 |
| einfacher Dienst (A1-4) | 11 | ,6 | ,6 | 61,4 |
| mittlerer Dienst (A5-8) | 38 | 2,0 | 2,0 | 63,3 |
| gehobener Dienst (A9-13) | 52 | 2,7 | 2,7 | 66,1 |
| höherer Dienst (A14-16) | 7 | ,4 | ,4 | 66,4 |
| kleiner Handwerks-, Landwirts-, Gewerbebetrieb | 30 | 1,6 | 1,6 | 68,0 |
| mittlerer Handwerks-, Landwirts-, Gewerbebetrieb | 15 | ,8 | ,8 | 68,8 |
| Akademiker(in), Freiberufler(in), Unternehmer(in) (größerer | 66 | 3,4 | 3,4 | 72,2 |
| sonstiges | 199 | 10,4 | 10,4 | 82,6 |
| nie erwerbstätig | 100 | 5,2 | 5,2 | 87,8 |
| fehlend | 234 | 12,2 | 12,2 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 45: Häufigkeitsverteilung des Prädiktors „Krankschreibung zu Beginn der Behandlung“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 1070 | 55,8 | 55,8 | 55,8 |
| ja | 595 | 31,0 | 31,0 | 86,8 |
| fehlend | 253 | 13,2 | 13,2 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 46: Häufigkeitsverteilung des Prädiktors „Einnahme von Schmerzmitteln“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozen-te |
|---|---|---|---|---|
| nie | 624 | 32,5 | 32,5 | 32,5 |
| an 1-3 Tagen pro Monat, Dosierung nach Bedarf | 432 | 22,5 | 22,5 | 55,1 |
| an 1-2 Tagen pro Woche, Dosierung nach Bedarf | 165 | 8,6 | 8,6 | 63,7 |
| an 3-7 Tagen pro Woche, Dosierung nach Bedarf | 114 | 5,9 | 5,9 | 69,6 |
| täglich, in fester Dosierung | 107 | 5,6 | 5,6 | 75,2 |
| fehlend | 476 | 24,8 | 24,8 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 47: Häufigkeitsverteilung des Prädiktors „Inanspruchnahme ambulanter Psychotherapie“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 674 | 35,1 | 35,1 | 35,1 |
| 1-5 Termine | 287 | 15,0 | 15,0 | 50,1 |
| 6-25 Termine | 465 | 24,2 | 24,2 | 74,3 |
| mehr als 25 Termine | 228 | 11,9 | 11,9 | 86,2 |
| fehlend | 264 | 13,8 | 13,8 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 48: Häufigkeitsverteilung des Prädiktors „Nikotinkonsum“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 1190 | 62,0 | 62,0 | 62,0 |
| ja | 621 | 32,4 | 32,4 | 94,4 |
| fehlend | 107 | 5,6 | 5,6 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 49: Häufigkeitsverteilung des Prädiktors „Arbeitsunfähigkeitszeiten für Berufstätige“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| 4 Wochen | 339 | 17,7 | 17,7 | 17,7 |
| 8 Wochen | 120 | 6,3 | 6,3 | 23,9 |
| 12 Wochen | 89 | 4,6 | 4,6 | 28,6 |
| 16 Wochen | 66 | 3,4 | 3,4 | 32,0 |
| 20 Wochen | 50 | 2,6 | 2,6 | 34,6 |
| 24 Wochen | 146 | 7,6 | 7,6 | 42,2 |
| über 24 Wochen | 137 | 7,1 | 7,1 | 49,4 |
| fehlend | 971 | 50,6 | 50,6 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 50: Häufigkeitsverteilung des Prädiktors „Einnahme von Antidepressiva“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nie | 814 | 42,4 | 42,4 | 42,4 |
| an 1-3 Tagen pro Monat, Dosierung nach Bedarf | 37 | 1,9 | 1,9 | 44,4 |
| an 1-2 Tagen pro Woche, Dosierung nach Bedarf | 18 | ,9 | ,9 | 45,3 |
| an 3-7 Tagen pro Woche, Dosierung nach Bedarf | 40 | 2,1 | 2,1 | 47,4 |
| täglich, in fester Dosierung | 582 | 30,3 | 30,3 | 77,7 |
| fehlend | 427 | 22,3 | 22,3 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 51: Häufigkeitsverteilung des Prädiktors „Geschlecht“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| weiblich | 1417 | 73,9 | 73,9 | 73,9 |
| männlich | 424 | 22,1 | 22,1 | 96,0 |
| fehlend | 77 | 4,0 | 4,0 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 52: Häufigkeitsverteilung des Prädiktors „Absicht einen Rentenantrag zu stellen“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 883 | 46,0 | 46,0 | 46,0 |
| ja | 84 | 4,4 | 4,4 | 50,4 |
| fehlend | 951 | 49,6 | 49,6 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 53: Häufigkeitsverteilung des Prädiktors „Antrag auf Erwerbsminderungsrente“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 1503 | 78,4 | 78,4 | 78,4 |
| ja | 86 | 4,5 | 4,5 | 82,8 |
| fehlend | 329 | 17,2 | 17,2 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 54: Häufigkeitsverteilung des Prädiktors „Einnahme von Schlafmitteln“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nie | 967 | 50,4 | 50,4 | 50,4 |
| an 1-3 Tagen pro Monat, Dosierung nach Bedarf | 100 | 5,2 | 5,2 | 55,6 |
| an 1-2 Tagen pro Woche, Dosierung nach Bedarf | 47 | 2,5 | 2,5 | 58,1 |
| an 3-7 Tagen pro Woche, Dosierung nach Bedarf | 52 | 2,7 | 2,7 | 60,8 |
| täglich, in fester Dosierung | 144 | 7,5 | 7,5 | 68,3 |
| fehlend | 608 | 31,7 | 31,7 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 55: Häufigkeitsverteilung des Prädiktors „Krankenhausaufenthalte“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 1364 | 71,1 | 71,1 | 71,1 |
| ja | 263 | 13,7 | 13,7 | 84,8 |
| fehlend | 291 | 15,2 | 15,2 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 56: Häufigkeitsverteilung des Prädiktors „Arbeitsunfähigkeitszeiten für Nicht-Berufstätige“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| 4 Wochen | 210 | 10,9 | 10,9 | 10,9 |
| 8 Wochen | 57 | 3,0 | 3,0 | 13,9 |
| 12 Wochen | 66 | 3,4 | 3,4 | 17,4 |
| 16 Wochen | 16 | ,8 | ,8 | 18,2 |
| 20 Wochen | 22 | 1,1 | 1,1 | 19,3 |
| 24 Wochen | 45 | 2,3 | 2,3 | 21,7 |
| über 24 Wochen | 34 | 1,8 | 1,8 | 23,5 |
| fehlend | 1468 | 76,5 | 76,5 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

### *16.3.2 Verteilungseigenschaften der fremd eingeschätzten Items auf nicht-metrischem Skalenniveau*

Tabelle 57: Häufigkeitsverteilung des Prädiktors „Arbeitsfähigkeit zu Beginn der Therapie“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 873 | 45,5 | 45,5 | 45,5 |
| ja | 909 | 47,4 | 47,4 | 92,9 |
| fehlend | 136 | 7,1 | 7,1 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 58: Häufigkeitsverteilung des Prädiktors „Schweregradzusatz der Erstdiagnose“

| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| | ,0 unauffällig | 13 | ,7 | ,7 | ,7 |
| | 1,0 leicht | 116 | 6,1 | 6,4 | 7,1 |
| | 2,0 mittel | 1093 | 57,6 | 60,5 | 67,7 |
| | 3,0 schwer | 584 | 30,8 | 32,3 | 100,0 |
| | Gesamt | 1806 | 95,2 | 100,0 | |
| Fehlend | System | 91 | 4,8 | | |
| Gesamt | | 1897 | 100,0 | | |

Tabelle 59: Häufigkeitsverteilung des Prädiktors „Rentenantrag“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 1688 | 88,0 | 88,0 | 88,0 |
| ja | 91 | 4,7 | 4,7 | 92,8 |
| fehlend | 139 | 7,2 | 7,2 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 60: Häufigkeitsverteilung des Prädiktors „Rentner/Rentnerin“

| | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|
| nein | 1564 | 81,5 | 81,5 | 81,5 |
| ja | 263 | 13,7 | 13,7 | 95,3 |
| fehlend | 91 | 4,7 | 4,7 | 100,0 |
| Gesamt | 1918 | 100,0 | 100,0 | |

Tabelle 61: Psychometrische Kennwerte des Prädiktors „Chronifizierung der Erkrankung“ (Ordinalskalenniveau)

| | | | Statistik | Standardfehler |
|---|---|---|---|---|
| Chronifizierung der Haupterkrankung (in Jahren) | Mittelwert | | 3,395 | ,0334 |
| | 95% Konfidenzintervall des Mittelwerts | Untergrenze | 3,329 | |
| | | Obergrenze | 3,460 | |
| | 5% getrimmtes Mittel | | 3,388 | |
| | Median | | 3,000 | |

| | | | |
|---|---|---|---|
| | Varianz | 2,068 | |
| | Standardabweichung | 1,4382 | |
| | Minimum | 1,0 | |
| | Maximum | 7,0 | |
| | Spannweite | 6,0 | |
| | Interquartilbereich | 3,0 | |
| | Schiefe | ,022 | ,057 |
| | Kurtosis | -,936 | ,114 |

## 16.4 Befragung im Rahmen der HoNOS-D Schulung

### *16.4.1 Fragebogen zur Erfassung der Praktikabilität (Auswahl)*




# Befragung im Rahmen der HoNOS-D Schulung

Arbeitsgruppe für Psychotherapie- und
Versorgungsforschung (AGPV)

Institut und Poliklinik für Medizinische Psychologie
Universitätsklinikum Hamburg-Eppendorf
Zentrum für Psychosoziale Medizin
Martinistr. 52, S 35
20246 Hamburg

Ansprechpartnerin:
Sylke Andreas
Diplom-Psychologin
Tel.: 040-42803-7705
Fax: 040-42803-4940
e-mail: sandreas@uke.uni-hamburg.de

**Liebe Schulungsteilnehmer, liebe Schulungsteilnehmerinnen,**

obgleich erste Hinweise zur Praktikabilität und routinemäßigen Verwendung der HoNOS aus internationalen Studien vorliegen, mangelt es bislang noch an Erfahrungen im deutschsprachigen Raum und bei bestimmten Patientengruppen. Daher möchten wir Sie im Folgenden um Ihre Unterstützung bitten, dazu mehr Kenntnisse zu erlangen.

Wir möchten Sie bitten den Fragebogen 2 Wochen nach der Schulung bzw. nach 5 Interviews mit dem Patienten auszufüllen und an die obige Adresse zurückzusenden.

| 1 | **Wie lange benötigen Sie zuletzt zum Ausfüllen der HoNOS-D?** |
|---|---|
| | ☐☐☐ Minuten |

| 2 | **Welche Items bereiteten Ihnen immer noch die größten Schwierigkeiten bei der Beantwortung?** |
|---|---|
| | Item ☐☐, Item ☐☐, Item ☐☐, Item ☐☐ |

| 3 | **Wie schätzen Sie die Verständlichkeit des Manuals (Glossars) ein?** |
|---|---|
| | sehr schlecht [0]- [1]- [2]- [3]- [4]- [5] ausgezeichnet |

| 4 | **Erhalten Sie alle nötigen Informationen für eine Einschätzung des HoNOS-D aus den routinemäßig durchgeführten Gesprächen (z.B. Aufnahme- oder Erstgespräche)?** |
|---|---|
| | ☐ nein |
| | ☞ Wenn nein, woher beziehen Sie dann Ihre Informationen weitest gehend? ___________________________ |
| | ☞ Wie hoch ist dann der zusätzliche Zeitaufwand (in Minuten)? ☐☐☐ Minuten |
| | ☐ ja |

| 5 | **Wie schätzen Sie die Praktikabilität des Instrumentes insgesamt ein?** |
|---|---|
| | sehr schlecht [0]- [1]- [2]- [3]- [4]- [5] ausgezeichnet |

| 6 Wie ist Ihre Meinung zu dem Instrument im Allgemeinen? | |
|---|---|
| positive Punkte: | 1. ___________________<br>2. ___________________<br>3. ___________________ |
| | |
| negative Punkte: | 1. ___________________<br>2. ___________________<br>3. ___________________ |

### *16.4.2 Ergänzende Auswertungen zur Praktikabilität der HoNOS-D*

Tabelle 62: Mittelwert und Standardabweichung zur Bearbeitungszeit im HoNOS-D

| | N | Minimum | Maximum | Mittelwert | Standardabweichung |
|---|---|---|---|---|---|
| Wie lange benötigen Sie zuletzt zum Ausfüllen der HoNOS-D? (Minuten) | 7 | 3,00 | 10,00 | 5,4286 | 2,14920 |
| Gültige Werte (Listenweise) | 7 | | | | |

Tabelle 63: Häufigkeitsverteilung der schwierigsten Items im HoNOS-D (Mehrfachnennungen) (n = 7 Beurteiler)

| | **Häufigkeit** | **Prozent** |
|---|---|---|
| Item 12 | 5 | 33,33 |
| Item 8 | 3 | 20,0 |
| Item 10 | 2 | 13,33 |
| Item 11 | 2 | 13,33 |
| Item 3 | 1 | 6,66 |
| Item 1 | 1 | 6,66 |
| Item 4 | 1 | 6,66 |
| Gesamt | 15 | 100 |

Tabelle 64: Mittelwert und Standardabweichungen zur Verständlichkeit des Glossars

| | N | Minimum | Maximum | Mittelwert | Standardabweichung |
|---|---|---|---|---|---|
| Wie schätzen Sie die Verständlichkeit des Manuals (Glossars) ein? | 9 | 1,00 | 4,00 | 3,1111 | ,92796 |
| Gültige Werte (Listenweise) | 9 | | | | |

Tabelle 65: Häufigkeitsverteilung des Items „Erhalten Sie alle nötigen Informationen aus dem routinemäßig durchgeführten Gesprächen?“

| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| | nein | 2 | 22,2 | 28,6 | 28,6 |
| | ja | 5 | 55,6 | 71,4 | 100,0 |
| | Gesamt | 7 | 77,8 | 100,0 | |
| Fehlend | System | 2 | 22,2 | | |
| Gesamt | | 9 | 100,0 | | |

## 16.5 Ergänzende Ergebnistabellen zur Überprüfung der generellen und spezifischen Eignung der HoNOS-D

### *16.5.1 Kalkulationsgrundlage der Finn-Koeffizienten (Interraterreliabilitäten)*

Der Finn-Koeffizient wurde nach folgender Formel berechnet:

$$r_{Finn} = 1 - \frac{MS_{beobachtet}}{(k^2 - 1) / 12}$$

k = Anzahl der Kategorien (= 6)

Im Folgenden folgt eine getrennte Darstellung der für die Berechnung zugrunde liegenden Werte für die beiden Teilstichproben. Dabei ist für $MS_{beobachtet}$ die Varianz innerhalb der Gruppen in die Formel einzusetzen.

Tabelle 66: Kennwerte der 1-faktoriellen Varianzanalyse der *Teilstichprobe 1*

| | | Quadrat-summe | df | Mittel der Quadrate | F | Signifikanz |
|---|---|---|---|---|---|---|
| honos1 | Zwischen den Gruppen | 28,941 | 33 | ,877 | 2,982 | ,001 |
| | **Innerhalb der Gruppen** | 10,000 | 34 | **,294** | | |
| | Gesamt | 38,941 | 67 | | | |
| honos2 | Zwischen den Gruppen | 78,191 | 33 | 2,369 | 1,627 | ,081 |
| | **Innerhalb der Gruppen** | 49,500 | 34 | **1,456** | | |
| | Gesamt | 127,691 | 67 | | | |
| honos3 | Zwischen den Gruppen | 51,657 | 33 | 1,565 | 5,740 | ,000 |
| | **Innerhalb der Gruppen** | 9,000 | 33 | **,273** | | |
| | Gesamt | 60,657 | 66 | | | |
| honos4 | Zwischen den Gruppen | 8,235 | 33 | ,250 | ,848 | ,681 |
| | **Innerhalb der Gruppen** | 10,000 | 34 | **,294** | | |
| | Gesamt | 18,235 | 67 | | | |
| honos5 | Zwischen den Gruppen | 139,132 | 33 | 4,216 | 1,124 | ,368 |
| | **Innerhalb der Gruppen** | 127,500 | 34 | **3,750** | | |
| | Gesamt | 266,632 | 67 | | | |
| honos6 | Zwischen den Gruppen | 248,721 | 33 | 7,537 | 1,964 | ,027 |
| | **Innerhalb der Gruppen** | 130,500 | 34 | **3,838** | | |
| | Gesamt | 379,221 | 67 | | | |
| honos7 | Zwischen den Gruppen | 12,235 | 33 | ,371 | ,970 | ,535 |
| | **Innerhalb der Gruppen** | 13,000 | 34 | **,382** | | |
| | Gesamt | 25,235 | 67 | | | |

Fortsetzung Tabelle 66

| | | | | | | |
|---|---|---|---|---|---|---|
| honos8 | Zwischen den Gruppen | 33,437 | 33 | 1,013 | ,904 | ,612 |
| | **Innerhalb der Gruppen** | 32,500 | 29 | **1,121** | | |
| | Gesamt | 65,937 | 62 | | | |
| honos9 | Zwischen den Gruppen | 26,721 | 33 | ,810 | 1,899 | ,034 |
| | **Innerhalb der Gruppen** | 14,500 | 34 | **,426** | | |
| | Gesamt | 41,221 | 67 | | | |
| honos10 | Zwischen den Gruppen | 49,142 | 33 | 1,489 | 3,389 | ,000 |
| | **Innerhalb der Gruppen** | 14,500 | 33 | **,439** | | |
| | Gesamt | 63,642 | 66 | | | |
| honos11 | Zwischen den Gruppen | 71,000 | 33 | 2,152 | 2,286 | ,009 |
| | **Innerhalb der Gruppen** | 32,000 | 34 | **,941** | | |
| | Gesamt | 103,000 | 67 | | | |
| honos12 | Zwischen den Gruppen | 85,857 | 33 | 2,602 | 2,602 | ,005 |
| | **Innerhalb der Gruppen** | 29,000 | 29 | **1,000** | | |
| | Gesamt | 114,857 | 62 | | | |

Anmerkungen: (fett = $MS_{beobachtet}$)

Tabelle 67: Kennwerte der 1-faktoriellen Varianzanalyse der *Teilstichprobe 2*

| | | Quadratsumme | df | Mittel der Quadrate | F | Signifikanz |
|---|---|---|---|---|---|---|
| honos1 | Zwischen den Gruppen | 13,471 | 33 | ,408 | 1,542 | ,107 |
| | **Innerhalb der Gruppen** | 9,000 | 34 | **,265** | | |
| | Gesamt | 22,471 | 67 | | | |
| honos2 | Zwischen den Gruppen | 5,132 | 33 | ,156 | 3,525 | ,000 |
| | **Innerhalb der Gruppen** | 1,500 | 34 | **,044** | | |
| | Gesamt | 6,632 | 67 | | | |
| honos3 | Zwischen den Gruppen | 3,779 | 33 | ,115 | ,865 | ,660 |
| | **Innerhalb der Gruppen** | 4,500 | 34 | **,132** | | |
| | Gesamt | 8,279 | 67 | | | |
| honos4 | Zwischen den Gruppen | 6,015 | 33 | ,182 | ,590 | ,933 |
| | **Innerhalb der Gruppen** | 10,500 | 34 | **,309** | | |
| | Gesamt | 16,515 | 67 | | | |

Fortsetzung Tabelle 67

| | | | | | | |
|---|---|---|---|---|---|---|
| honos5 | Zwischen den Gruppen | 28,529 | 33 | ,865 | 1,960 | ,027 |
| | **Innerhalb der Gruppen** | 15,000 | 34 | **,441** | | |
| | Gesamt | 43,529 | 67 | | | |
| honos6 | Zwischen den Gruppen | 5,471 | 33 | ,166 | ,939 | ,571 |
| | **Innerhalb der Gruppen** | 6,000 | 34 | **,176** | | |
| | Gesamt | 11,471 | 67 | | | |
| honos7 | Zwischen den Gruppen | 21,191 | 33 | ,642 | 2,569 | ,004 |
| | **Innerhalb der Gruppen** | 8,500 | 34 | **,250** | | |
| | Gesamt | 29,691 | 67 | | | |
| honos8 | Zwischen den Gruppen | 20,765 | 33 | ,629 | 2,139 | ,015 |
| | **Innerhalb der Gruppen** | 10,000 | 34 | **,294** | | |
| | Gesamt | 30,765 | 67 | | | |
| honos9 | Zwischen den Gruppen | 14,250 | 33 | ,432 | 1,727 | ,059 |
| | **Innerhalb der Gruppen** | 8,500 | 34 | **,250** | | |
| | Gesamt | 22,750 | 67 | | | |
| honos10 | Zwischen den Gruppen | 17,309 | 33 | ,525 | 2,098 | ,017 |
| | **Innerhalb der Gruppen** | 8,500 | 34 | **,250** | | |
| | Gesamt | 25,809 | 67 | | | |
| honos11 | Zwischen den Gruppen | 18,941 | 33 | ,574 | 1,301 | ,225 |
| | **Innerhalb der Gruppen** | 15,000 | 34 | **,441** | | |
| | Gesamt | 33,941 | 67 | | | |
| honos12 | Zwischen den Gruppen | 12,103 | 33 | ,367 | ,851 | ,675 |
| | **Innerhalb der Gruppen** | 12,500 | 29 | **,431** | | |
| | Gesamt | 24,603 | 62 | | | |

### *16.5.2 Die Entscheidungsbäume der zehn Zufallsstichproben*

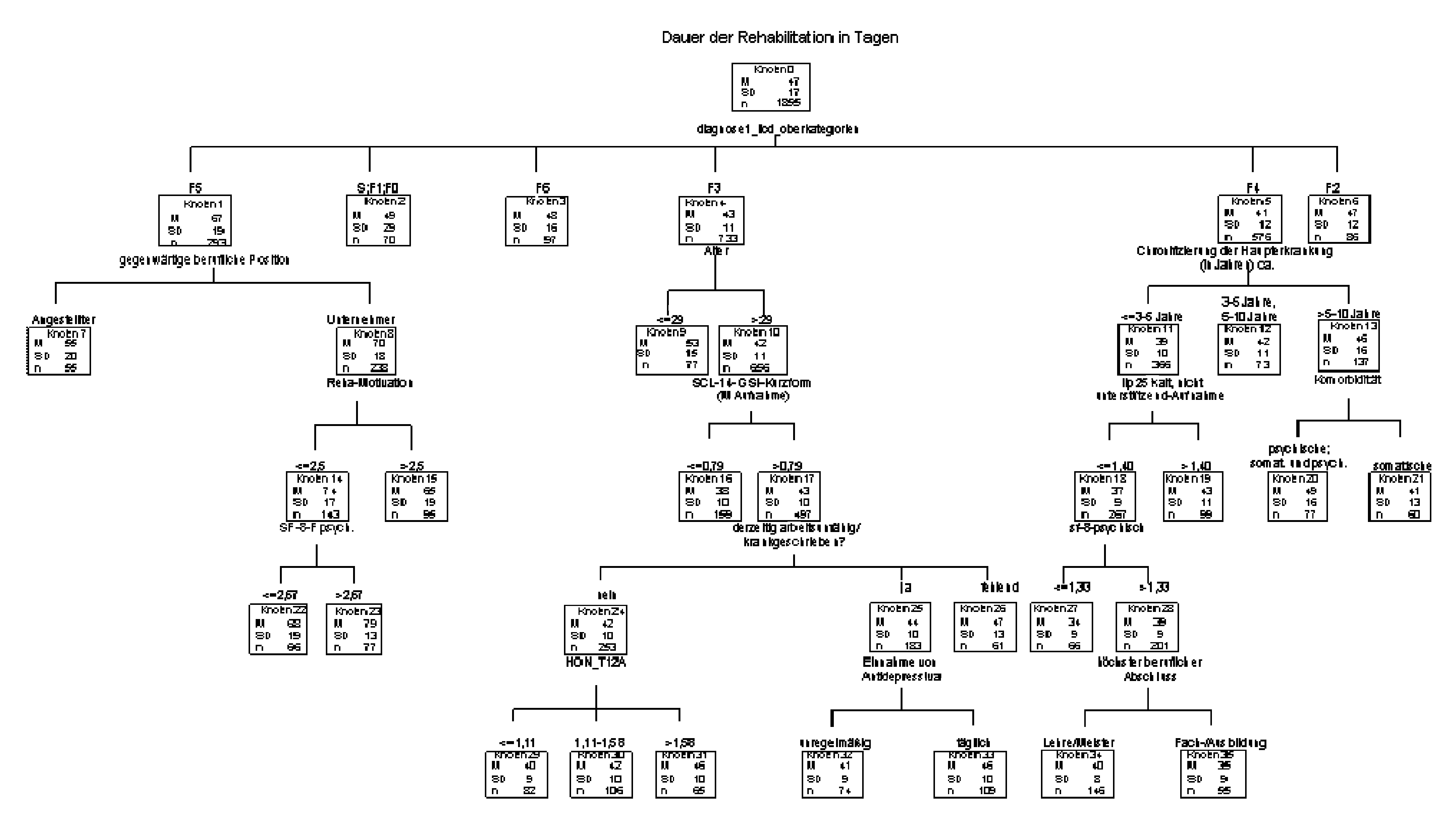



Abbildung 50: Entscheidungsbaum der ersten Zufallsstichprobe (N = 1855) mit 35.50% Varianzaufklärung.

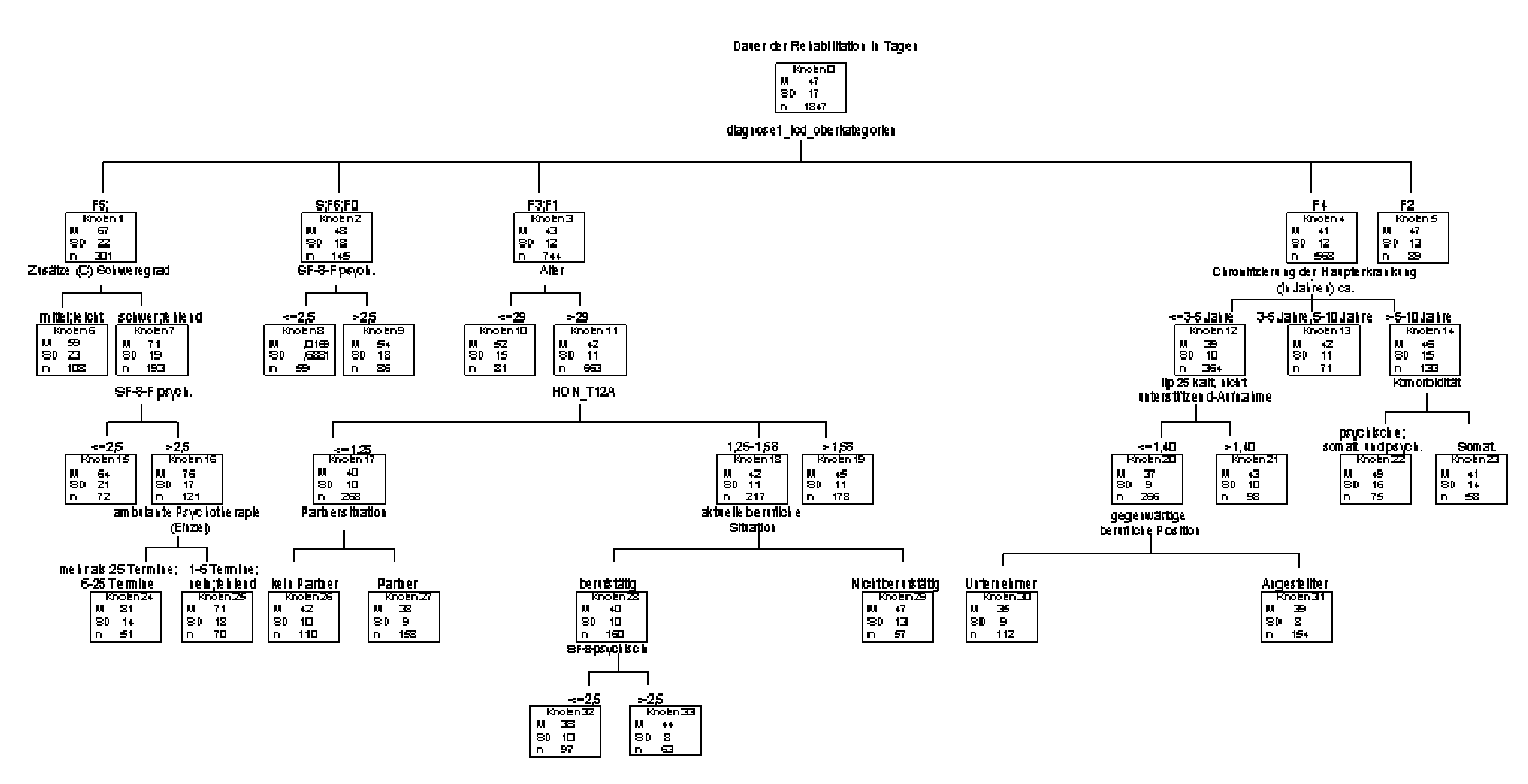



Abbildung 51: Entscheidungsbaum der zweiten Zufallsstichprobe (N = 1847) mit 34.83% Varianzaufklärung.

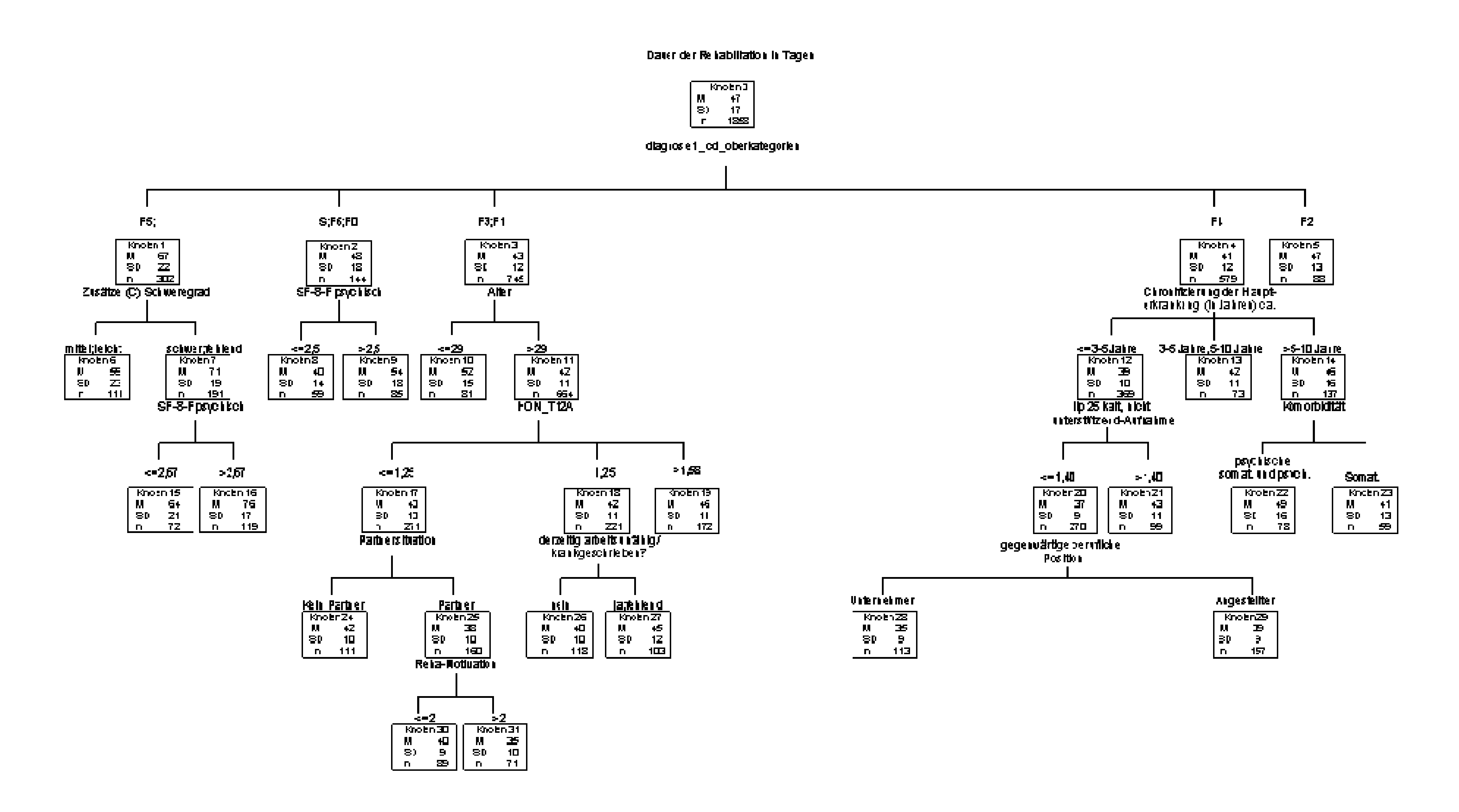



Abbildung 52: Entscheidungsbaum der dritten Zufallsstichprobe (N = 1858) mit 35.71% Varianzaufklärung.

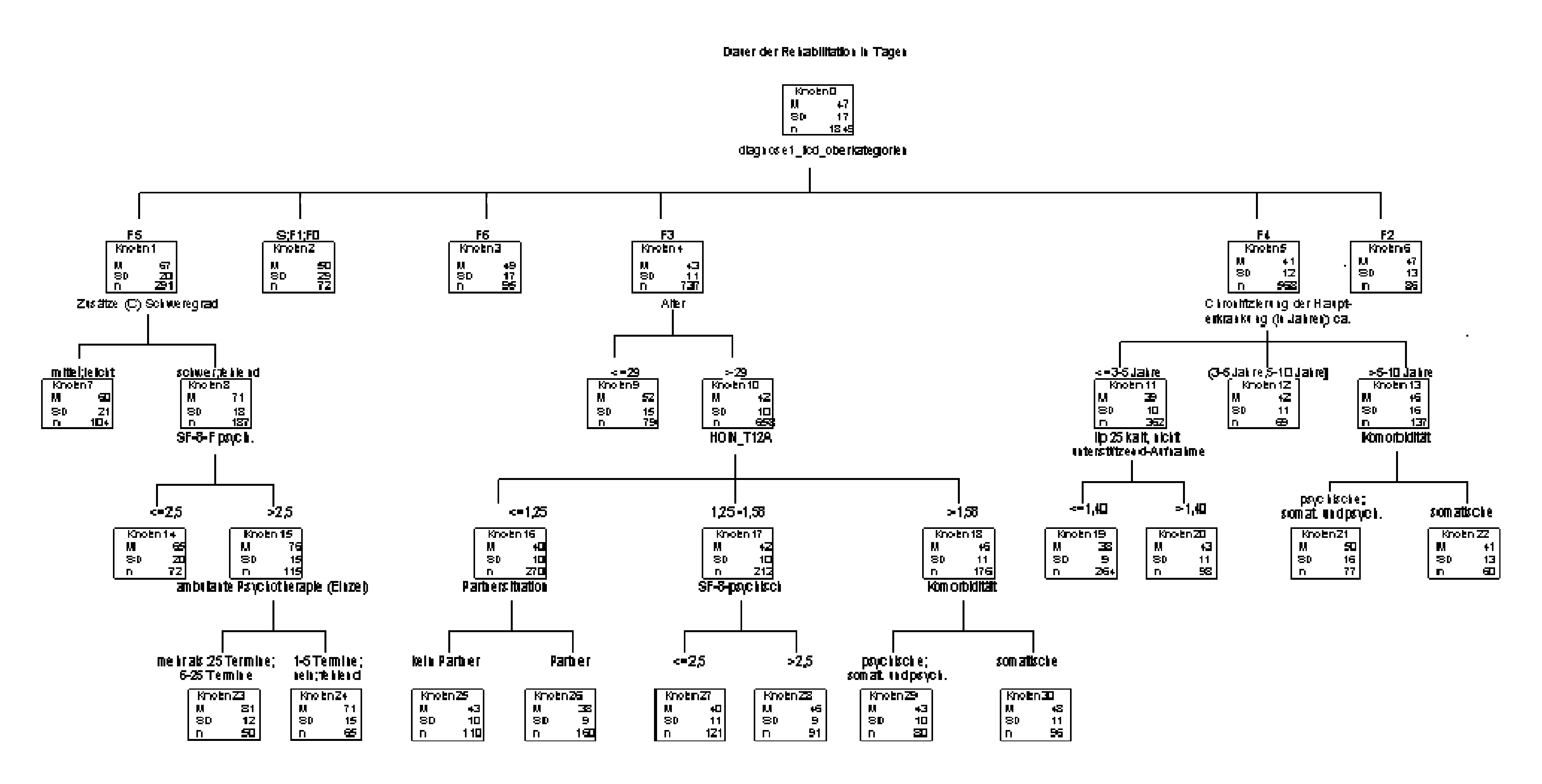



Abbildung 53: Entscheidungsbaum der vierten Zufallsstichprobe (N = 1849) mit 33.92% Varianzaufklärung.

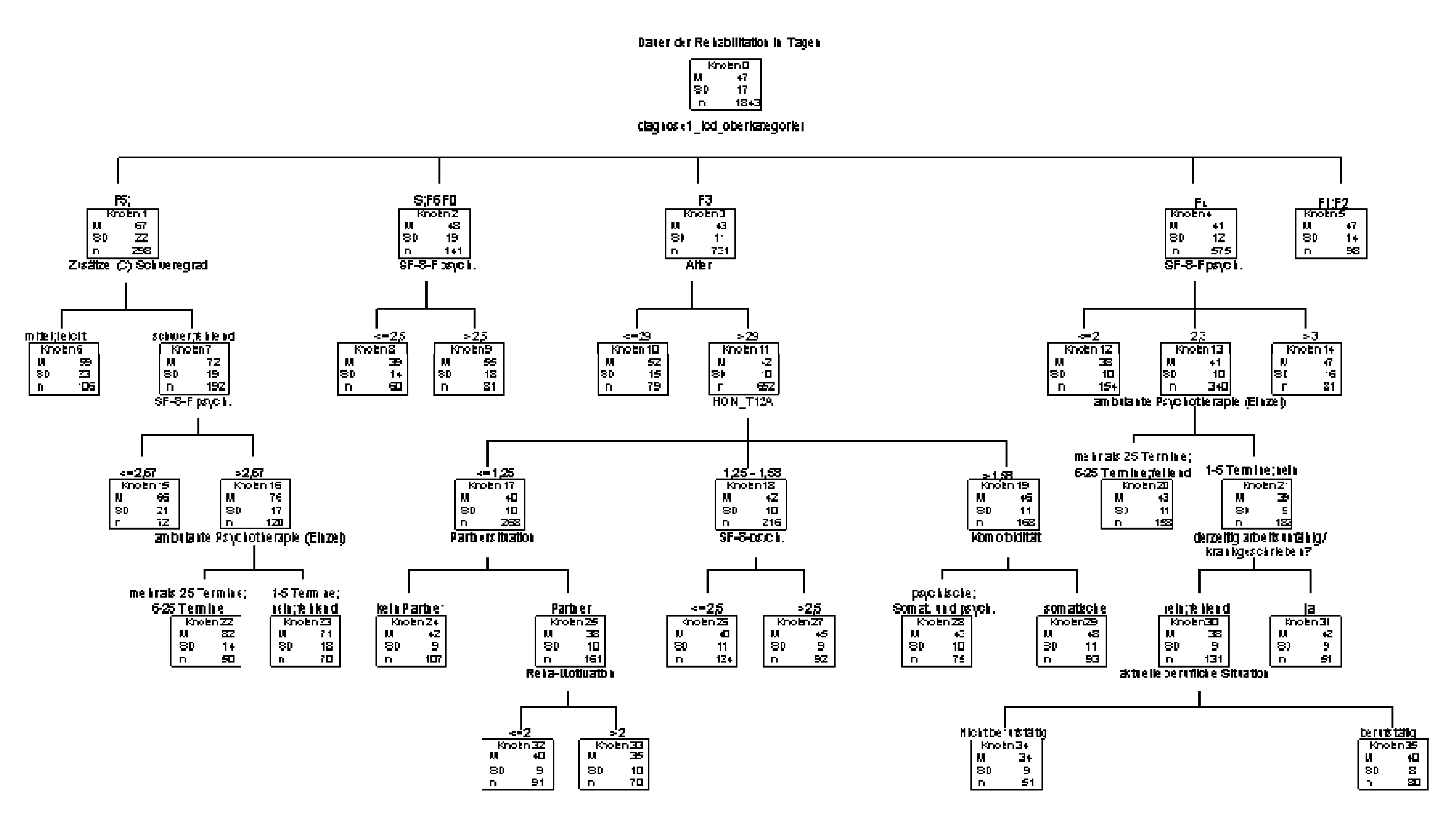



Abbildung 54: Entscheidungsbaum der fünften Zufallsstichprobe (N = 1843) mit 37.11% Varianzaufklärung.

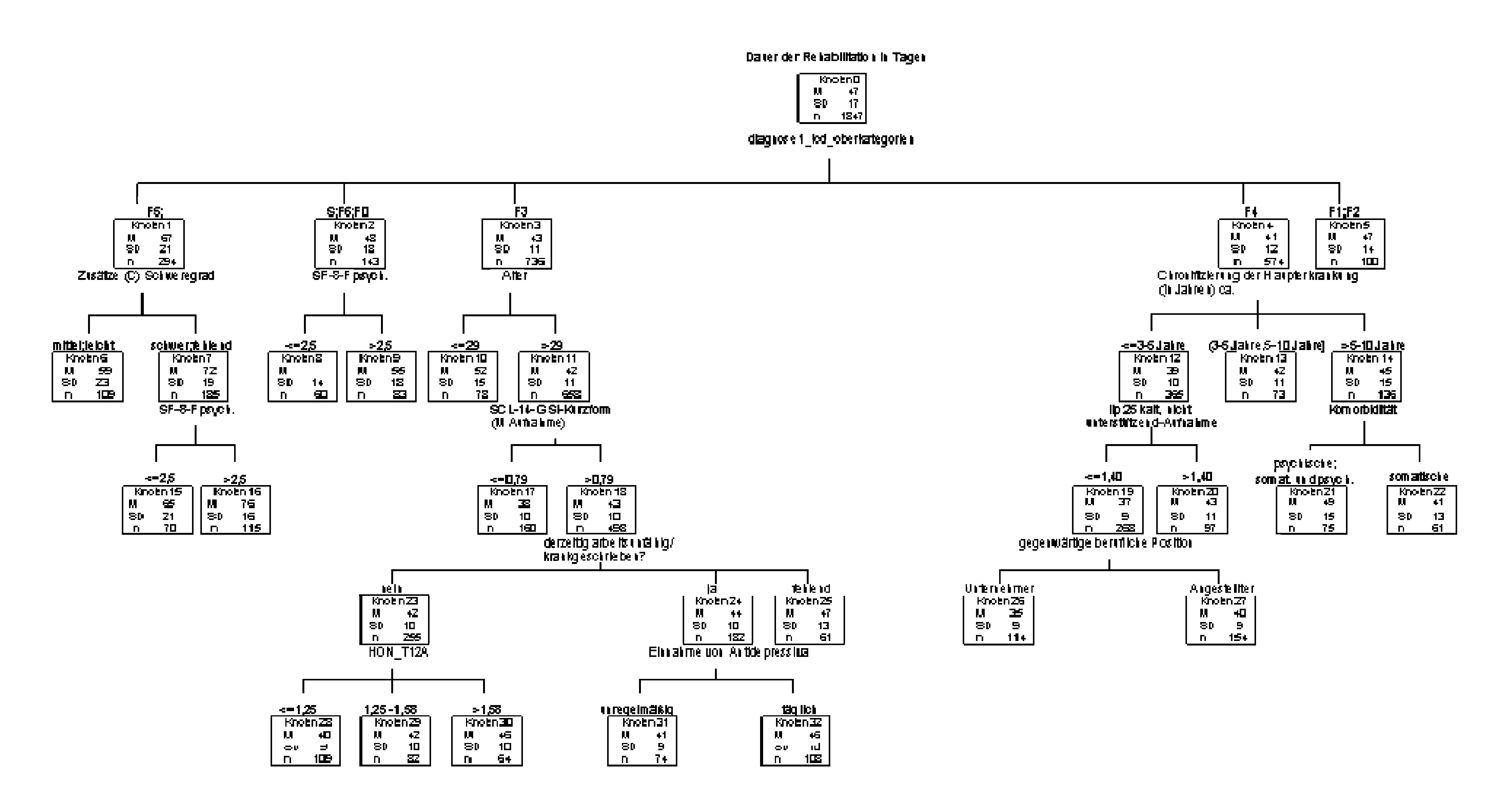



Abbildung 55: Entscheidungsbaum der sechsten Zufallsstichprobe (N = 1847) mit 36.16% Varianzaufklärung.

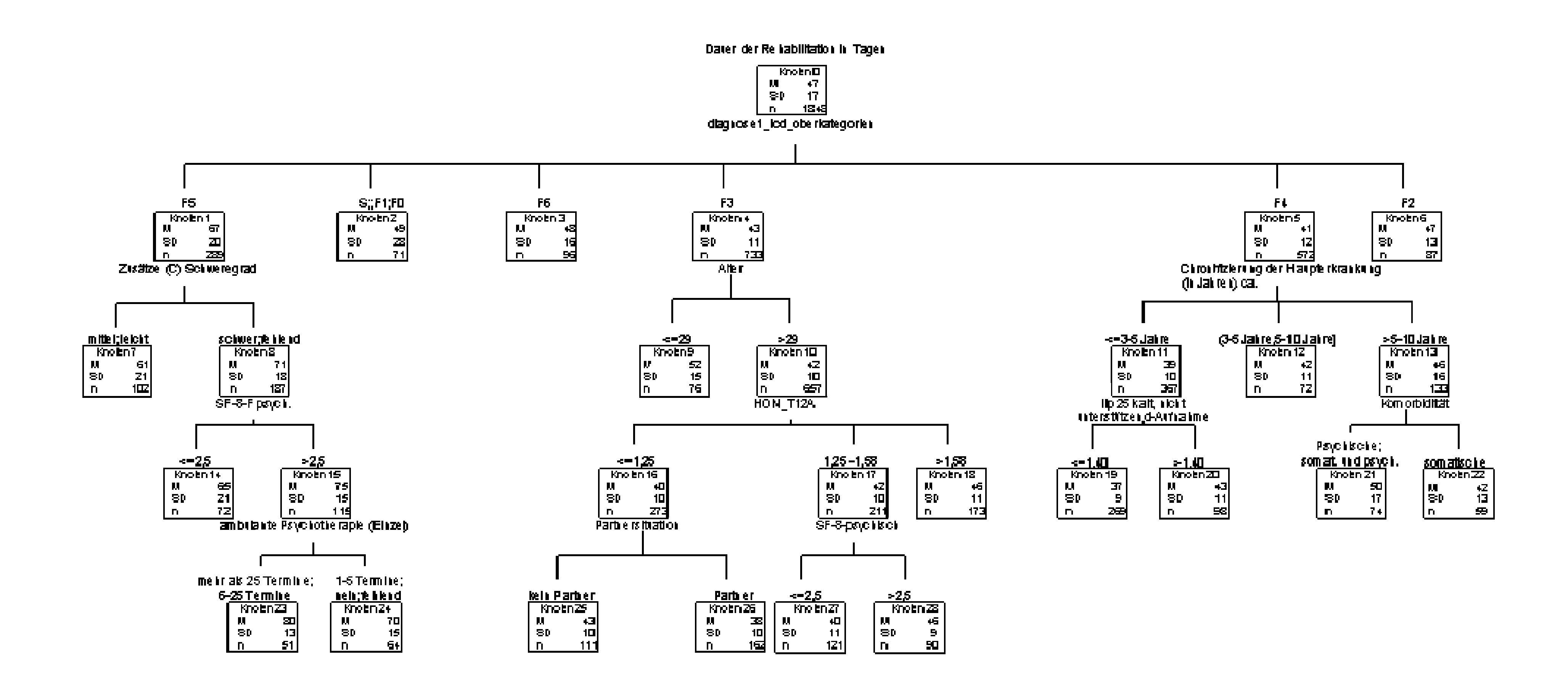


Abbildung 56: Entscheidungsbaum der siebten Zufallsstichprobe (N = 1848) mit 33.95% Varianzaufklärung.

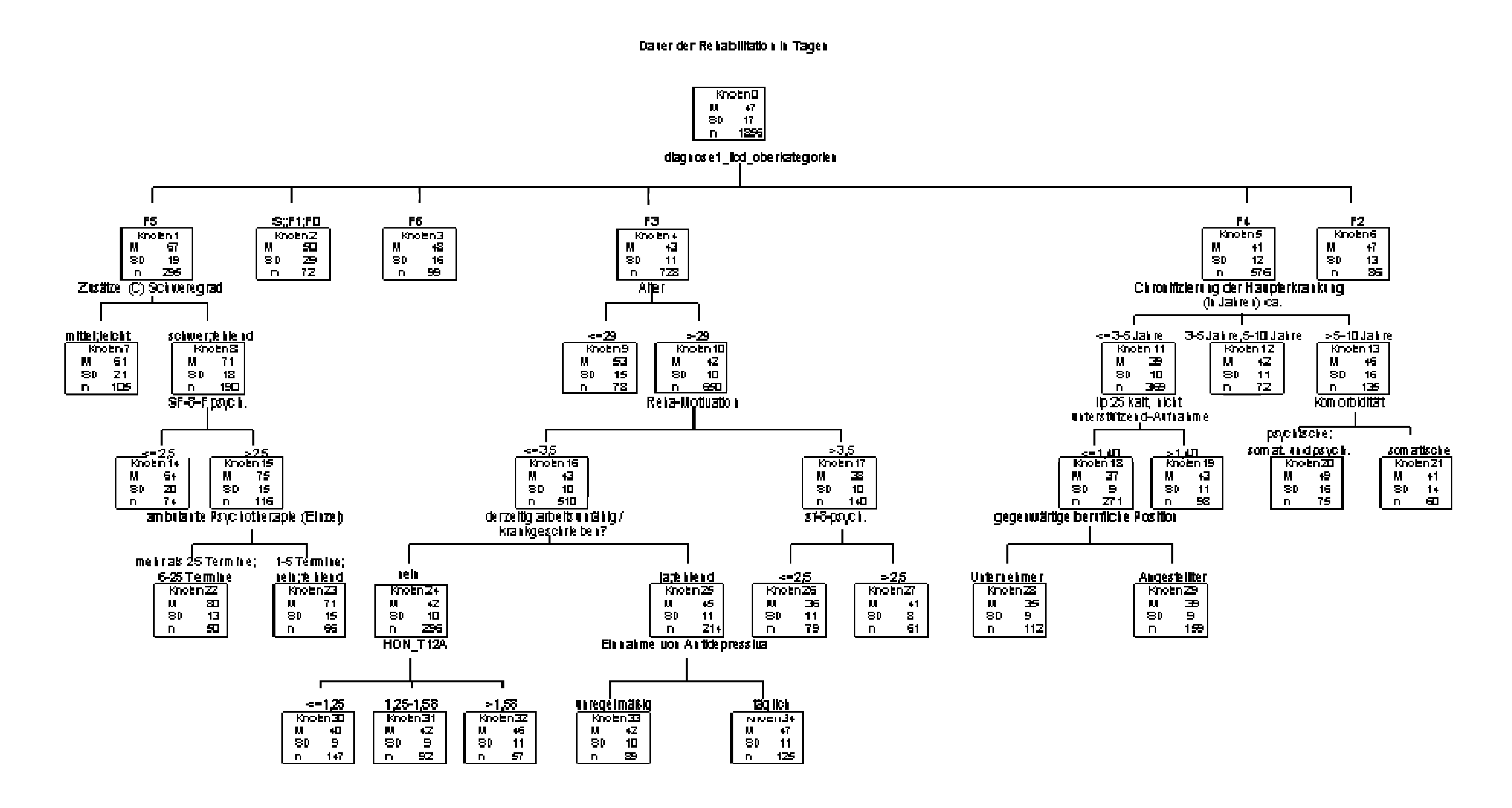



Abbildung 57: Entscheidungsbaum der achten Zufallsstichprobe (N = 1856) mit 34.26% Varianzaufklärung.

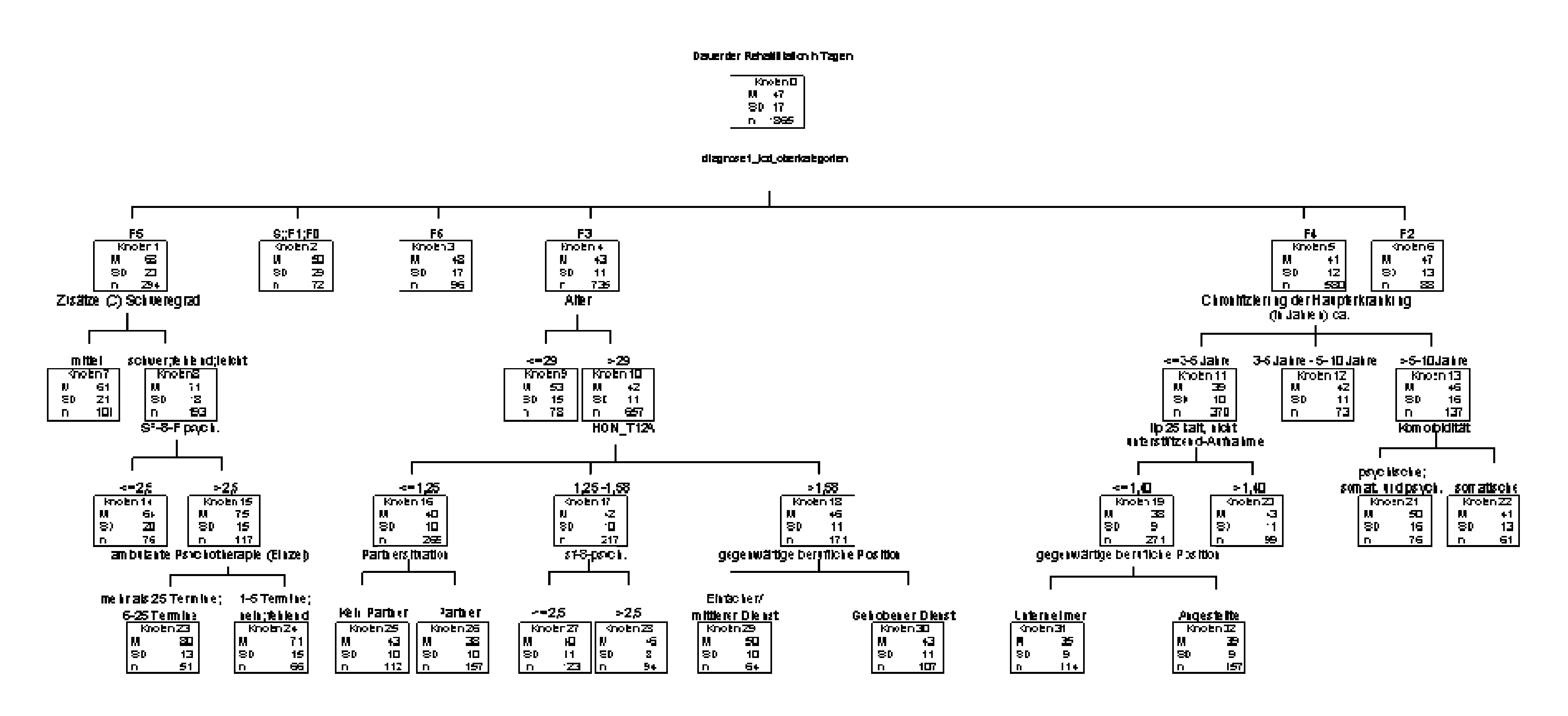



Abbildung 58: Entscheidungsbaum der neunten Zufallsstichprobe (N = 1885) mit 34.83% Varianzaufklärung.

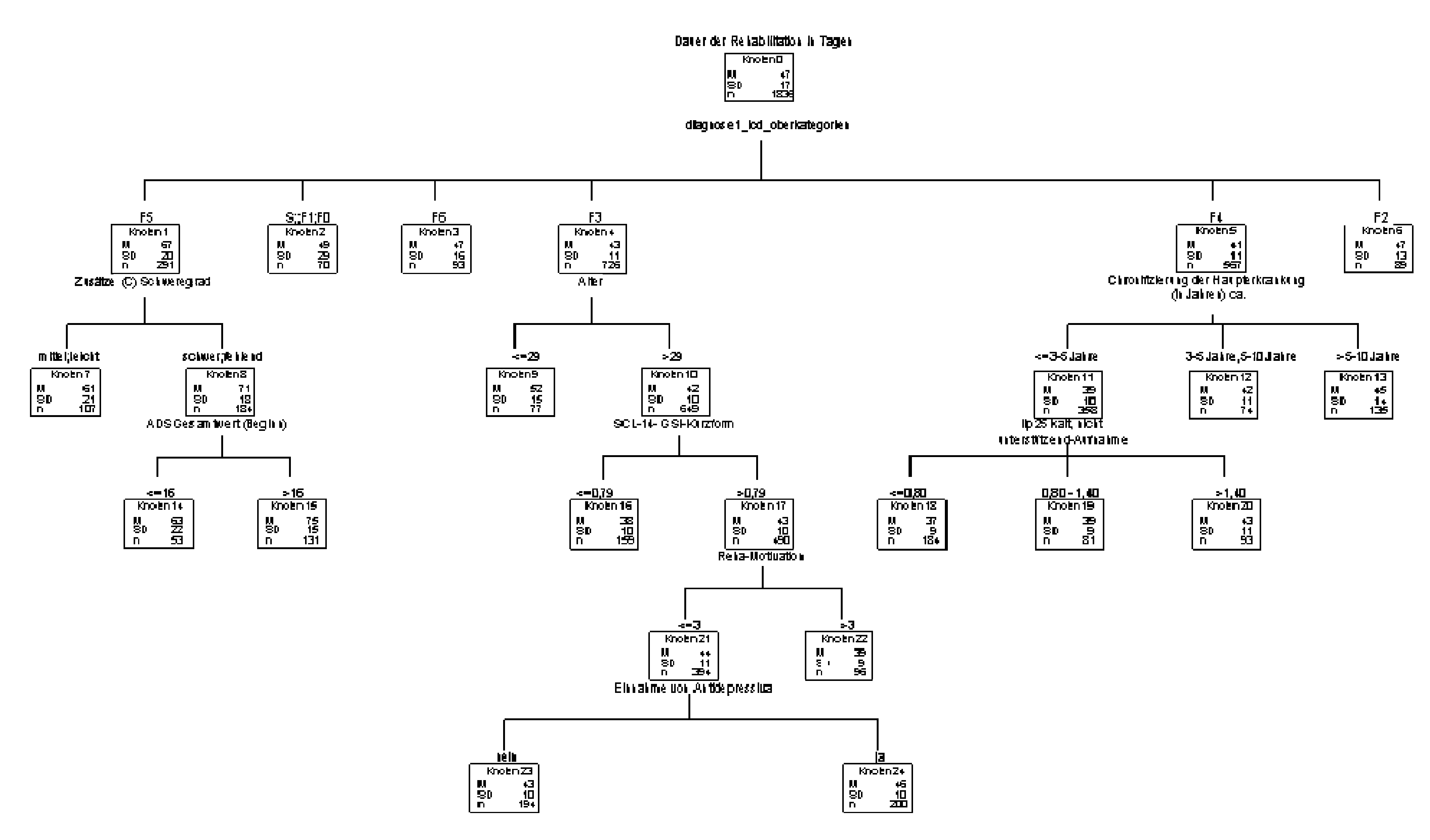



Abbildung 59: Entscheidungsbaum der zehnten Zufallsstichprobe (N = 1836) mit 33.84% Varianzaufklärung.

## 16.6 Vorschlag für eine Modifikation der HoNOS-D

## Beurteilerinstruktion zum Rating im HoNOS-D

*Sehr geehrte Behandler und Behandlerinnen,*

anbei möchten wir Ihnen gern **ergänzend** zum Glossar und zum Rating-Sheet des HoNOS-D einen **kurzen Beurteilerleitfaden** zur Verfügung stellen, mit den wichtigsten Informationen, die Sie für eine Beurteilung im HoNOS-D benötigen.

1. Zur Beurteilung können alle **Informationsquellen** (z.B. auch Angaben von Pflegekräften genutzt werden.
2. Der Beurteilungszeitraum umfasst **14 Tage**: Es sollte also eingeschätzt werden, wie schwer der Patient in den 12 Items der HoNOS-D in den letzten 14 Tagen beeinträchtigt war.
3. Jedes Item wird auf einer **5-Punkt-Skala** eingeschätzt (von 0 bis 4).
4. Es sollte **vermieden** werden, die „9" (d.h. „unbekannt") zu vergeben.
5. Für die Items 1 bis 9 gilt: Bei jedem Item soll das **schlimmste** Problem eingeschätzt werden, um ein Abbild des gegenwärtigen Zustandes zu erhalten.
6. Items 10 bis 12 dagegen benötigen eine eher **generelle** Einschätzung.
7. **Vermeiden** Sie **Überlappungen** zwischen den Items: Der Inhalt eines Items sollte nur **1x** eingeschätzt werden:

   *Zum Beispiel*: Jemand war in einen Kampf verwickelt (die Schwere der Aggression wird bei Item 1 eingeschätzt), er ist auch betrunken (Schwere des Alkoholproblems bei Item 3 einschätzen) und leidet unter einer physischen Verletzung (Schwere der Verletzung bei Item 5 einschätzen).

8. Items sollten in **numerischer** Reihenfolge (also von 1 bis 12) eingeschätzt werden.
9. Bei den **Beispielen**, die im Glossar hinter jedem Ankerpunkt verzeichnet sind, handelt es sich um Beispiele, die auftreten *können*; *es müssen nicht* alle aufgeführten Beispiele bei einem Patienten vorliegen, für die Vergabe einer „1" (d.h. „klinisch unbedeutendes Problem) bei *Item 1* reicht es z.B. aus, wenn der Patient reizbar ist.
10. **Item 2** ist mit „Absichtlicher Selbstverletzung" tituliert, bezieht sich aber auch auf Suizidgedanken und –versuche.
11. **Item 8** enthält zehn weitere klinisch u.U. relevante Problembereiche:
    - Hier möchten wir Sie gern bitten, zunächst jeden dieser **zehn Problembereiche**, die unter Item 8 zusammengefasst sind, **einzeln einzuschätzen**
    - Anschließend möchten wir Sie gern bitten, **dasjenige Problem auszuwählen**, in dem der **Patient am schwersten belastet ist,** und den entsprechenden Buchstaben in das Kästchen eintragen

Nun sollen noch einige Anmerkungen zu den **Items 10 bis 12** gegeben werden.

1. **Item 10 „Probleme mit alltäglichen Aktivitäten"**: Beurteilen Sie bitte das Niveau der aktuellen Leistungsfähigkeit
   - Das Grundniveau beinhaltet selbstversorgende Aktivitäten, wie z.B. Essen, Waschen, Anziehen, zur Toilette gehen => Ist die Durchführung schwer beeinträchtigt, dann geben Sie bitte eine „3" oder „4"
   - Das komplexe Niveau beinhaltet den Gebrauch von höheren Fähigkeiten (z.B. Umgehen mit Geld, Einkaufen, Kinderversorgung) => Ist die Durchführung angemessen, dann geben Sie bitte eine „0" oder eine „1"
2. **Item 11 „Probleme durch die Wohnbedingungen"**
   - In die Einschätzung dieses Items gehören: z.B. Sind die Grundbedürfnisse in der häuslichen Umgebung abgedeckt (Wärme, Licht)? => Wenn nicht, dann geben Sie bitte eine „4"
   - Berücksichtigen Sie hier auch die Beziehungen des Patienten zu Freunden und das Ausmaß, in dem das Umfeld des Patienten geeignet ist, seine Motivation zu erhöhen und Verhaltensmuster zu verbessern
3. **Item 12 „Probleme durch die Bedingungen im Beruf und in der Freizeit"**
   - die Einschätzung des Items 12 basiert auf einer Gesamteinschätzung des Ausmaßes, in welchem die tägliche Umgebung die besten Fähigkeiten des Patienten hervorbringt
   - Wenn der Grad der Unabhängigkeit des Patienten nicht durch seine Umgebung eingeschränkt ist (bzw. durch Bedingungen im Beruf und im Alltag) => Dann geben Sie bitte eine „0"

## HoNOS-D Rating zu Behandlungsbeginn

Beurteilen Sie für jedes der 12 Items das jeweils **schwerste** Problem, das während der **letzten 14 Tage** aufgetreten ist.

**Legende**: Tragen Sie für jedes Item 0 - 4 oder 9 in die Kästchen ein.

| | | | |
|---|---|---|---|
| **0** | kein Problem | **1** | klinisch unbedeutendes Problem ohne Handlungsbedarf |
| **2** | leichtes Problem, aber eindeutig vorhanden | **3** | eher schweres Problem |
| **4** | schweres bis sehr schweres Problem | **9** | nicht bekannt / nicht anwendbar |

1) Überaktives, aggressives, Unruhe stiftendes oder agitiertes Verhalten ☐

2) Absichtliche Selbstverletzung ☐

3) Problematischer Alkoholkonsum oder Drogenkonsum ☐

4) Kognitive Probleme ☐

5) Probleme in Zusammenhang mit körperlicher Erkrankung oder Behinderung ☐

6) Probleme in Zusammenhang mit Halluzinationen und Wahnvorstellungen ☐

7) Gedrückte Stimmung ☐

8) Andere psychische und verhaltensbezogene Probleme

| | | | |
|---|---|---|---|
| *A Phobisch* | ☐ | *B Angst* | ☐ |
| *C Zwangsgedanken/ -handlungen* | ☐ | *D Psychische Belastung / Anspannung* | ☐ |
| *E Dissoziativ* | ☐ | *F Somatoform* | ☐ |
| *G Essen* | ☐ | *H Schlaf* | ☐ |
| *I Sexuell* | ☐ | *J Andere (spezifizieren)* | ☐ |

______________________

*Bitte geben Sie nun noch den **Bereich (A-J)** an, in dem das schwerste Problem aufgetreten ist, indem Sie hier den **Buchstaben** eintragen.* ☐ **A-J**

9) Probleme mit Beziehungen ☐

10) Probleme mit alltäglichen Aktivitäten ☐

11) Probleme durch die Wohnbedingungen ☐

12) Probleme durch die Bedingungen im Beruf und im Alltag ☐